

INTERNATIONAL ENCYCLOPEDIA  
OF THE  
SOCIAL SCIENCES

8777  
4985



S.G.B.R.T. West Bengal

Date .....

Acc. No. 4985 .....



International  
Encyclopedia of the  
SOCIAL  
SCIENCES

Volumes 13 and 14  
Complete and Unabridged

## Associate Editors

Heinz Eulau, *Political Science*  
Lloyd A. Fallers, *Anthropology*  
William H. Kruskal, *Statistics*  
Gardner Lindzey, *Psychology*  
Albert Rees, *Economics*  
Albert J. Reiss, Jr., *Sociology*  
Edward Shils, *Social Thought*

## Special Editors

Elinor G. Barber, *Biographies*  
John G. Darley, *Applied Psychology*  
Bert F. Hoselitz, *Economic Development*  
Clifford T. Morgan, *Experimental Psychology*  
Robert H. Strotz, *Econometrics*

## Editorial Staff

Marjorie A. Bassett, *Economics*  
P. G. Bock, *Political Science*  
Robert M. Coen, *Econometrics*  
J. M. B. Edwards, *Sociology*  
David S. Gochman, *Psychology*  
George Lowy, *Bibliographies*  
Judith M. Tanur, *Statistics*  
Judith M. Treistman, *Anthropology*

## Alvin Johnson

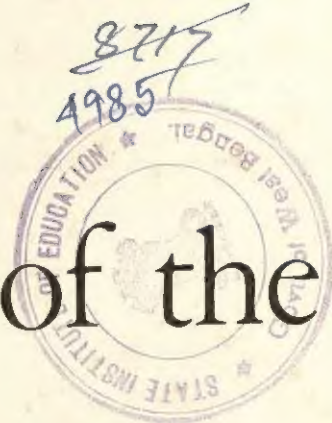
HONORARY EDITOR

## W. Allen Wallis

CHAIRMAN, EDITORIAL ADVISORY BOARD



# International Encyclopedia of the SOCIAL SCIENCES



DAVID L. SILLS EDITOR

VOLUME 13

The Macmillan Company & The Free Press, New York  
COLLIER-MACMILLAN PUBLISHERS, LONDON

1985

COPYRIGHT © 1968 BY CROWELL COLLIER AND MACMILLAN, INC.

ALL RIGHTS RESERVED UNDER THE INTERNATIONAL COPYRIGHT UNION,  
THE INTER-AMERICAN COPYRIGHT UNION, AND UNDER THE PAN-AMERICAN  
COPYRIGHT CONVENTIONS.

NO PART OF THIS BOOK MAY BE REPRODUCED OR TRANSMITTED IN  
ANY FORM OR BY ANY MEANS, ELECTRONIC OR MECHANICAL, INCLUDING  
PHOTOCOPYING, RECORDING, OR BY ANY INFORMATION STORAGE AND  
RETRIEVAL SYSTEM, WITHOUT PERMISSION IN WRITING FROM  
CROWELL COLLIER AND MACMILLAN, INC.

MANUFACTURED IN THE UNITED STATES OF AMERICA

REPRINT EDITION 1972



International  
Encyclopedia of the  
SOCIAL  
SCIENCES

# P

[ CONTINUED ]

## PSYCHOANALYSIS

*The articles under this entry provide an introduction to psychoanalytic theory and methods of therapy. The biography of FREUD should also be consulted; it includes the historical background of Freud's thought, his major contributions to psychoanalysis, and a critique of his theories. Some major implications of psychoanalysis for behavior are discussed in ANXIETY; BODY IMAGE; CONFLICT, article on PSYCHOLOGICAL ASPECTS; DEFENSE MECHANISMS; EMOTION; IDENTITY, PSYCHOSOCIAL; NATIONAL CHARACTER; PERSONALITY, POLITICAL; SELF CONCEPT; SOCIALIZATION, article on PSYCHOLOGICAL ASPECTS; SUICIDE. Discussions that are relevant to an evaluation of the libido concept are contained in AFFECTION; AGGRESSION, article on PSYCHOLOGICAL ASPECTS; DRIVES; INSTINCT; MOTIVATION. The relevance of psychoanalysis for the study of mental disorders is discussed in CHARACTER DISORDERS; DEPRESSIVE DISORDERS; HYSTERIA; NEUROSIS; OBSESSIVE-COMPULSIVE DISORDERS; PARANOID REACTIONS; PHOBIAS; PSYCHOPATHIC PERSONALITY; PSYCHOSIS; PSYCHOSOMATIC ILLNESS; SCHIZOPHRENIA. Psychoanalytic interpretations of the products of human behavior are contained in DREAMS; FANTASY; LITERATURE, article on THE PSYCHOLOGY OF LITERATURE; RELIGION, article on PSYCHOLOGICAL STUDY. Methods of studying phenomena of interest to psychoanalysis are discussed in DREAMS; HYPNOSIS; INTERVIEWING, especially the article on THERAPEUTIC INTERVIEWING; PROJECTIVE METHODS. Different and often contrasting views of personality are found in ANALYTICAL PSYCHOLOGY; FIELD THEORY; GESTALT THEORY; INDIVIDUAL PSYCHOLOGY; LEARNING THEORY; PERSONALITY; PERSONALITY: CONTEMPORARY*

*VIEWPOINTS; PHENOMENOLOGY; PSYCHOLOGY, article on EXISTENTIAL PSYCHOLOGY; THINKING, article on COGNITIVE ORGANIZATION AND PROCESSES. Different conceptions of mental disorders and their treatment are described in CLINICAL PSYCHOLOGY; MENTAL DISORDERS; MENTAL DISORDERS, TREATMENT OF; MENTAL HEALTH; PSYCHIATRY; PSYCHOLOGY, article on EXISTENTIAL PSYCHOLOGY. Biographies of important figures (in addition to Freud) relevant either to Freud's work or the psychoanalytic movement are ABRAHAM; ADLER; ALEXANDER; BLEULER; CHARCOT; FERENCZI; HORNEY; JONES; JUNG; KLEIN; KRIS; RANK; RAPAPORT; REICH; RÖHEIM; RORSCHACH; SULLIVAN.*

- I. CLASSICAL THEORY
- II. EGO PSYCHOLOGY
- III. THERAPEUTIC METHODS
- IV. EXPERIMENTAL STUDIES

Charles Brenner  
George S. Klein  
Rudolph M. Loewenstein  
Ernest R. Hilgard

### I CLASSICAL THEORY

The term "psychoanalysis" is commonly used in three different senses: a form of treatment for mental illness, a method for investigating the workings of the mind (the psychoanalytic method), and a branch of psychological or behavioral science. The term was coined by Sigmund Freud, who devised, developed, and applied the method over a period of fifty years and who is responsible for the major part of the theoretical formulations called psychoanalysis.

Psychoanalysis began as a method of treating mentally ill patients, and the theoretical formulations which Freud first made on the basis of his clinical experience concerned psychopathology. However, in the course of his work Freud came to recognize that there are close similarities between



psychopathology and certain aspects of normal mental functioning. Thus by 1900 it was clear that psychoanalysis had definite contributions to make to the psychology of dreams, of jokes, and of various slips and errors of everyday life, which Freud proposed to call normal or everyday psychopathology. These were momentous discoveries which proved to be fruitful as well. Even more momentous was the discovery of the vital importance of childhood sexuality in both normal and pathological mental development, a discovery which was contained in a monograph published in 1905 and entitled *Three Essays on the Theory of Sexuality*.

The subsequent development of psychoanalytic theory, based on the application of the psychoanalytic method to an increasingly large number of cases, seems to justify the claim that it constitutes by far the most important contribution to the psychology of man that has been made up to this time. Psychoanalysis has been described epigrammatically as human nature viewed as conflict. And indeed it appears to be the case that many of the most important aspects of human mental development and functioning are intimately related to conflicts which have their origins in childhood and in particular in the sexual conflicts of childhood. This statement is as true of character traits and attributes which we class as normal as it is of those aspects of behavior which we class as symptoms of mental illness.

This, then, is the most important and the most unexpected contribution of psychoanalysis to normal psychology: the vital significance of instinctual conflict in normal as well as pathological mental functioning and development. Closely allied to this is the importance in mental life of unconscious mental phenomena, that is, of mental processes of which the subject himself has no conscious knowledge. In fact, the conflicts just mentioned, which play such an important part in mental life, are largely unconscious in later childhood and adult mental life. It may be noted parenthetically that in psychoanalytic terminology at present "unconscious" usually means "accessible to consciousness only with difficulty, or not at all," while "preconscious" means "readily accessible to consciousness, though not conscious at the moment." However, current psychoanalytic usage is not wholly consistent with respect to the meanings of "unconscious" and "preconscious," since their meanings have varied somewhat in the course of the development of Freud's theories concerning the mental apparatus.

A third contribution of psychoanalysis to normal

psychology is its demonstration of the continuity or determinism of mental life. If one depends on simple introspection for one's knowledge of what goes on in the mind, there appear to be many gaps and discontinuities in the current of mental life. It often happens that an idea comes to consciousness which bears no apparent connection with what one was consciously thinking a moment before. Similarly, behavior may bear no apparent connection to conscious volition or conscious thoughts. If, however, one is able to apply the psychoanalytic method, one is in a position to adduce evidence for the existence of unconscious mental processes that fill in the gaps and discontinuities which appear to be present in mental life if one judges only by the data of conscious introspection.

Finally, in addition to demonstrating the causal relationship between the present and the immediate past in mental life, the application of the psychoanalytic method has made it possible to establish relationships between current modes of thought and behavior, on the one hand, and various crucially important experiences and conflicts of early childhood, on the other. The emphasis on the relationship between the present and the more remote past in mental life, which is so characteristic of psychoanalytic psychology, is often referred to as the genetic point of view.

In summary, we may repeat that psychoanalysis, which began as a theory of psychopathology, has gradually developed into an important part of general psychology. The application of the psychoanalytic method has contributed more to our understanding of the *important* aspects of human mental functioning and development than has any other method of observation or experiment.

### Assumptions and method

As a behavioral science, psychoanalysis is primarily concerned with the inner life of man. It assumes that man's thoughts and ideas, his plans and fantasies, and his fears and daydreams are legitimate objects of study. It also assumes, as does biological science in general, that man is an animal. For the psychoanalyst, psychology is a branch of biology. The importance of somatic or organic influences on mental functioning is always borne in mind, particularly with relation to the instinctual drives, which, according to psychoanalytic theory, are assumed to be the basic motivational forces in mental life.

**Free association.** The method of psychoanalysis is often referred to, rather inaccurately, as free association. The subject, who usually lies on a couch



facing away from the observer or analyst, is instructed to communicate freely to the analyst, without reserve or editing, whatever thoughts occur to him, as well as whatever physical or bodily sensations intrude upon his consciousness. The analyst may, and often does, suggest that the patient associate to some particular idea or experience that he has communicated in this way. For example, he may suggest that the subject associate to one or another part of a dream which he has related. In addition, he may intervene with what are usually called interpretations in psychoanalytic terminology; that is, he may point out to the subject evidences of conflict, often unconscious conflict, within the subject's mind. At other times, the analyst may attempt to reconstruct the nature of childhood experiences and fantasies which are important determinants of present behavior or mental activity. [See DREAMS.]

From these very brief examples it should be clear that the term "free association" is far from a satisfactory description of the psychoanalytic method, although it does point to certain essential elements of the method. Indeed the assumption that underlies the application of the psychoanalytic method is that if conscious control of one's thoughts is given up, the direction of one's thoughts, far from being free, will then be determined even more clearly than before by other motivational forces within the mind which are not subject to conscious volition and which are in many instances unconscious. To be more precise, the application of the psychoanalytic method reveals more clearly the presence and the consequences of inner conflict, conflict of which the subject himself often has little or no knowledge. Since, by the nature of things, the recognition of such conflict within oneself gives rise to shame, guilt, anxiety, or other unpleasant emotions, there is generally a strong resistance to revealing the presence of such conflict to another individual or even to oneself. In human mental life the most important things of all are often those which one least wants to tell. A conscious determination not to be influenced by one's reluctance to speak of certain things is only partially helpful. The reluctance itself is in fact lifelong and to a large extent unconscious. It is for this reason that the satisfactory application of the psychoanalytic method is generally possible only in a therapeutic situation. Fortunately for our scientific curiosity, though unfortunately for each of us personally, there is enough that is neurotic even in the most normal of us so that any psychoanalysis soon becomes a therapeutic one, even though it may have been originally under-

taken merely out of scientific curiosity or as part of a course of training.

**Behavioral observation.** In addition to listening to what the subject tells him of what is going on in his mind, the analyst observes the subject's behavior during the entire time when it is accessible to his observation, neglecting no aspect or detail of it. It should be obvious that evidence of strong emotions, for example, tears, laughter, blushing, or anger, will be as important to the analyst as is the verbal content of what his subject tells him. Just as important are the subject's tone of voice, his incidental gestures, and his behavior both before he lies down on the couch and after he rises from it.

In general, the psychoanalytic method may be described as a descriptive method as opposed to a quantitative one, a naturalistic approach as opposed to an experimental one. It is not apparent a priori that the psychoanalytic method is a reliable one. Its reliability is attested by the regularity with which observations can be confirmed or repeated by the same analyst on the same subject, by the same analyst on different subjects, and by different analysts on different subjects.

**Other methods of study.** Although the application of the psychoanalytic method has been and remains the chief source of data as far as psychoanalysis is concerned, it must not be supposed that it is the only one. In recent years in particular, the method of direct observation of children over a period of months or years of their development has come to occupy an increasingly important place in the field of psychoanalytic research. Other observations of human behavior also have contributed data of various sorts, in particular the observation of the behavior of mentally ill patients in clinical settings other than the psychoanalytic treatment situation. Finally, mention should be made of the importance of the study of the various products of man's imagination. These include such examples of collective or folk creativity as myths, legends, religion, fairy tales, and, to a certain degree, the entire range of man's social institutions or systems. Equally important are the products of individual creativity, such as works of literature or the visual arts, and closely allied to this is the study of individual history or biography. In the psychoanalytic literature, studies of this sort are generally referred to as applied psychoanalysis. The explanatory value of psychoanalytic theory in these various areas of human activity affords a valuable, if subsidiary, additional confirmation of the validity of the primary observations which are obtainable by the ap-



plication of the psychoanalytic method—and of the theories which are primarily based upon those observations. [See FANTASY; LITERATURE, *article on THE PSYCHOLOGY OF LITERATURE.*]

### The instinctual drives

Psychoanalytic theories of the mind are often divided into two major parts, the first of which comprises those aspects of the theory that have to do with the instinctual drives and the second those aspects that have to do with what Freud called the mental apparatus. Although this division has the disadvantage of being excessively schematic, it does have the advantage of somewhat simplifying the problem of presenting an extremely complex subject. We shall therefore adhere to it, and begin our presentation with a discussion of the instinctual drives.

As the term "drive" implies, the instinctual drives are assumed to be forces within the mind which impel it to activity. They are the driving forces of mental life. In addition to their impulsive quality, they have certain other characteristics, namely, a *source*, an *object*, and an *aim*. The source of an instinctual drive is assumed to be some bodily part or bodily process. The object is the person (or thing) which is necessary for the satisfaction of the drive. The aim is the particular activity or action by means of which satisfaction is achieved. The capacity of an instinctual drive to impel the mind to activity is accounted for by assuming that in mental life an instinctual drive possesses *energy*. This energy is assumed to be quantitatively variable, to be increased at times of instinctual tension or need and to be discharged in the process of instinctual gratification. The quantity of instinctual energy which is associated with a particular memory or idea is referred to as the *cathexis* of that mental element. [See DRIVES; INSTINCT; MOTIVATION.]

**Erotic and aggressive drives.** Instinctual drives are divided into two major groups or types. The first is called the erotic or libidinal drive and the second the aggressive drive. The relationship of the erotic drive to bodily parts and processes is much clearer than is that of the aggressive drive, and the erotic drive in general has been more thoroughly studied and is better understood than the aggressive one. The principal sources of erotic drive tension or need in the body are referred to as *erotogenic zones*. The most important of these appear to be the mouth and related structures, the anus, and the genitalia. In the normal course of development it appears that the oral manifestations of the erotic drive are of particularly great importance

during the first year of life, the anal ones during the next year and a half, and the phallic ones thereafter. However, the sexual life of childhood is not dominated exclusively by any one erotogenic zone or by any single aim, as is normally the case in adult life after adolescence. On the contrary, the sexual wishes and satisfactions of childhood are characterized by their diversity and manifold quality rather than by uniformity. What is normally characteristic for childhood sexuality, therefore, would be called *perverse* in adult life.

Since the somatic sources, if any, of aggression (the aggressive drive) are less clear than those of the erotic drive, it is not possible to describe phases of development of the aggressive drive in childhood in the same way that one can describe the erotic drive. It appears that there is a close relationship between manifestations of the aggressive and the erotic drives throughout life, however—a fact which is explained by assuming a degree of fusion between the two drives. It follows therefore that the oral, anal, and phallic phases of childhood sexuality are characterized by aggressive as well as by erotic wishes, fantasies, and behavior, so that it is proper to speak of oral, anal, or phallic aggressive impulses.

An important characteristic of the aggressive drive is its differentiation with respect to its object. Freud was led to the assumption of a special aggressive drive in the first place by his recognition of the very great importance in mental life of self-injurious and self-destructive tendencies. It is clearly of utmost importance to any individual whether his aggressive impulses are directed toward himself or toward the mental representations of other persons and objects. It must not be assumed however that self-directed aggression is to be considered always pathological. On the contrary, it plays an important, indeed an essential role in normal mental life, for example, in the functioning of the superego, as well as a very important one in abnormal mental life. [See AGGRESSION, *article on PSYCHOLOGICAL ASPECTS.*]

The instinctual drives are of particular importance with respect to the emotional or affective life. In general, an intensification of instinctual need, or an increase in instinctual energy striving for discharge, is experienced as a source of pain and discomfort. A discharge of instinctual energy, that is, the gratification of an instinctual wish or need, which leads to a diminution in tension, is experienced as pleasurable.

**Evaluation.** The following summary statement may be of some value in helping to avoid certain misunderstandings concerning the current status

of the psychoanalytic theory of the instinctual drives. The fact is that some aspects of that theory are considered to be much better validated than others. For example, it seems certain that any qualified observer who is in a position to apply the psychoanalytic method to the study of human mental life must agree that the sexual life of man begins long before puberty. The fact of infantile sexuality (childhood sexuality) appears to be incontrovertible. Equal validity would appear to accrue to the importance of sexual and aggressive wishes, as well as conflicts concerning those wishes, in the whole of mental development and functioning throughout life. Validating evidence is much less adequate concerning such aspects of instinctual-drive theory as the concepts of instinctual energy, cathexis, fusion, and related concepts of a quantitative nature. We may add, parenthetically, that in the psychoanalytic literature the quantitative aspect of instinctual-drive theory is frequently referred to as its economic aspect.

### The psychic apparatus

The primary purpose of the psychic apparatus is assumed to be the regulation, control, and discharge of instinctual energy. It tends to function according to the *pleasure principle*, that is, in such a way as to achieve the pleasure of gratification which results from a discharge of instinctual energies. Formulated more generally, it may be said that the avoidance of discomfort or pain, which is associated with instinctual tension, and the achievement of pleasure, which is associated with gratification, is a basic tendency or regulatory principle of the psychic apparatus.

Psychoanalytic theories concerning the psychic apparatus are firmly based on the fact that the adult organization and functioning of the mind is the result of a period of gradual evolution and development during the course of the many years of childhood. In the last analysis, this is due to biological factors, that is, to the fact that the human infant is born into the world in a largely undeveloped state, both with respect to mental and physical functioning. As a result, during the first several years of its life, it is dependent on its parents for its very survival, as well as for the gratification of its instinctual needs. The offspring of no other mammal is so dependent on its parents for such a long period of time as is the human, and this fact is of profound importance in the mental life and development of man.

**Structure of the psychic apparatus.** In its mature form or organization the psychic apparatus is divided into three groups of related functions.

*The id.* That group of psychic functions which represents the instinctual drives in mental life and plays the role of stimulating or impelling the psychic apparatus to activity is called the id. The id is the part of the psychic apparatus which is in closest or most direct communication with the somatic processes of an instinctual nature, and it has often been referred to for that reason as the vital or somatic stratum of the psychic apparatus.

*The ego.* The ego is that part of the psychic apparatus which comprises the group of functions that have to do with the individual's relationship to his environment. Ego functions include such gradually developing capacities of the mind as sensory perception, memory, thought, and voluntary motor control. These functions cannot appear before the neuromuscular apparatus which is a prerequisite for them has matured. The development of ego functions, therefore, is a result of interaction between maturation of the nervous system and various experiential factors. What is particularly characteristic of the psychoanalytic theory of the psychic apparatus, however, is the following additional consideration. Ego functions appear to serve primarily the purpose of gratification of instinctual drives. They are, so to say, the executors of the wishes or urges which are the mental representations of the drives. There is thus from the very beginning the most intimate connection between the functioning and, indeed, the development of the mental capacity to cope with the external environment, on the one hand, and the instinctual needs and wishes of each individual, on the other.

*The superego.* The third group of functions that constitute the psychic apparatus is known as the superego. These functions develop into a more or less coherent unity rather late in childhood, probably not before the sixth or seventh year. In general they are that group of mental functions which have to do with the moral prohibitions and ideals of the individual.

**Conflict.** It is of particular importance to bear the following fact in mind in considering the psychoanalytic theory of the psychic apparatus: The functional groupings of id, ego, and superego which are assumed to constitute the mental apparatus are sharply differentiated from one another only in situations of inner conflict. Indeed, the very divisions reflect the characteristic varieties of conflict within the mind as they can be observed by means of the psychoanalytic method. Thus, one frequent type of conflict is between an instinctual wish on the one hand and those aspects of mental activity on the other which have to do with the



individual's relationship to those about him, that is, between a wish of the id and various ego functions. Characteristically, in later life at least, the moral prohibitions (superego) are arrayed on the side of the ego against the id. Other frequently observed conflicts are between the so-called ego functions, on the one hand, and the self-punitive or self-destructive aspect of the moral demands of the mind (superego), on the other. Thus, in these typical situations of conflict, one can distinguish properly between groups of functions that have to do with instinctual demands, those that have to do with the external world, and those that have to do with moral prohibitions and ideals. In other situations, however, where there is no such inner conflict, there is no sharp line of division or cleavage that enables us to distinguish clearly between or among the three systems, or structures as they are often called, of the psychic apparatus. If ego functions are acting as executants of the drives without conflict, one can no longer distinguish clearly what is ego from what is id. All one sees is a harmoniously functioning whole. The same is true when superego demands or prohibitions are in harmony with the instinctual needs or wishes of the individual and with the demands of his environment. [See CONFLICT, article on PSYCHOLOGICAL ASPECTS.]

### The id

The id is that part of the mental apparatus which functions to impel the mind toward the gratification of the instinctual drives. The impulsive aspect of mental life is that which is most directly associated with the id.

Since the controlling and adaptive aspects of mental functioning, which are subsumed under the concept of the ego, are only developed gradually during the course of infancy and childhood, the id gives the impression of being the most archaic as well as the most infantile part of the psychic apparatus. In many senses it is correct to say that it represents the mind of the child that lives on in the mind of the grown man. Those processes which we assume constitute the id proceed according to the pleasure principle with little or no concern for the restrictions or demands either of external reality or of morality. These are the aspects of mental functioning which poets and philosophers in the past have referred to as demonic.

The functioning of the id is characterized by an insistent or imperious demand for gratification, that is, for discharge of mental energy. This mode of mental functioning is called the *primary process* in psychoanalytic terminology. As we have just

noted, its essential attribute is a persistent striving for prompt and complete discharge of instinctual energy, that is, for prompt and full gratification. In keeping with this basic tendency, there is a propensity to substitute one or another instinctual object or instinctual aim for the primary one. Such substitution is called *displacement*, the reference being to the displacement of instinctual energy from one to another mental element. In addition to displacement, there may occur convergence of various instinctual cathexes on one particular aim or object which, because of its availability, or for other reasons, permits more ready and complete discharge of energy. Such a phenomenon is called *condensation*. These two related phenomena of displacement and condensation are typical and characteristic of mental phenomena which proceed according to the primary-process mode of discharge.

### The ego

The ego is that group of mental functions which has to do primarily with the relation of the individual to the external environment. These functions operate as the executants of the instinctual drives and may thus be considered to mediate between the inner needs of the individual and his environment.

Some of the functions that are grouped together as the ego appear to be directly related to the form and functioning of the nervous system, as, for example, sensory perception and the control of voluntary muscular activity, while others, such as the ability to distinguish between subjective and objective reality (originally called *reality principle*), are more complex and more obviously related to experiential or learning processes in the course of development.

From a subjective, psychological point of view, the single most important object in an infant's environment is his own body. His experiences of gratification and distress, and consequently his wishes and memories, are all centered about this unique object. Other objects of the environment, for example, his mother, although they may be of the greatest importance as far as his physical survival is concerned, do not compare with his own body in the first months and years of life as far as their psychological significance to him is concerned. These relationships are reflected in psychoanalytic theory by the statements, first, that the ego is first and foremost a body ego, and, second, that the infant's instinctual orientation is at first largely a narcissistic one, that is, at first largely directed toward himself.

The proper development and organization of

various ego functions as the infant grows appear to be intimately dependent upon a satisfactory balance between the instinctual gratifications which are supplied by the persons in his environment and the frustrations imposed by them. Proper mothering and, later, proper fathering as well are essential to proper development of the psychic apparatus and, more broadly, of the personality as a whole. The particular details of this complex interrelationship have been the object of many clinical and developmental studies. In psychoanalytic terminology, the most general formulation to express this relationship is that satisfactory object relationships promote the proper development of ego functions, while unsatisfactory ones impede their development.

**Identification and ego development.** Among the more obvious reasons why satisfactory object relationships are of importance in the development of ego functions is the fact that many ego functions develop by means of identification with the persons in the child's or infant's environment. This is most clearly apparent with respect to the acquisition of language, an accomplishment which is of the utmost importance for the individual's capacity to adapt satisfactorily to his environment. However, it is true to an important extent for other ego functions as well. In psychoanalytic terminology, identification may be defined as the tendency to become like an object of one's environment in one or more ways. It is also used to describe the end result of such a tendency. Identification appears to be intimately and necessarily connected with the objects of one's instinctual wishes. The child identifies or tends to identify with those persons in his environment with whom he has a strong emotional tie. He wishes and tends to be or become like those whom he loves and hates most intensely. [See IDENTITY, PSYCHOSOCIAL; IMITATION.]

**Delayed gratification.** The most general characteristic of the ego, one which is of fundamental importance in the development of most, if not all, of its other functions, is the capacity to postpone instinctual gratification, that is, to delay the discharge of instinctual energy. In psychoanalytic terminology, instinctual energy whose discharge is thus delayed is referred to as bound energy, in contradistinction to free energy. The introduction of delay into the process of instinctual discharge or gratification represents a modification of the pleasure principle and a deviation from the tendency to immediate discharge of instinctual energy, which, as noted above, is characteristic of the primary process. It is an essential capacity of the ego to impose delay or postponement of the gratification of an instinctual wish. This capacity

permits a more efficient exploitation of the environment in the service of the individual's search for gratification. The fact that the flux of mental energies which is characteristic of ego functions differs from that which characterizes the functioning of the id led Freud to the formulation that whereas the id functions according to the primary process, the ego functions according to the *secondary process*.

**Defense mechanisms and anxiety.** In addition to its capacity to delay the discharge of instinctual energies, the ego also includes among its functions those mental mechanisms which are referred to as defenses. It will be recalled from the introduction that the study of the consequences and nature of intrapsychic conflict occupies a central position in psychoanalytic theory. It is impossible to understand these conflicts without a thorough discussion of the mode of operation of the ego's defenses, which is part and parcel of the psychoanalytic theory of anxiety. Because of their centrally important position in mental functioning, therefore, the topics of ego defenses and of anxiety will be treated rather more extensively than the other ego functions.

Freud based his theory of anxiety on the assumption that whenever the mental apparatus is subjected to an influx of stimuli which exceeds the capacity of the apparatus either to discharge or to bind the energies of the stimuli, a traumatic state develops. The affect which accompanies this traumatic state is anxiety. In early infancy and childhood, the capacity of the mental apparatus to bind stimuli and to discharge them is normally much less than it is later on in life. For this reason traumatic states accompanied by anxiety appear relatively frequently in early life. In particular, such traumatic states develop in the infant when an instinctual need arises in the absence of the object, usually the mother, necessary for the gratification of that need, that is, for the discharge of the instinctual energy connected with it. As the infant grows and as its ego functions develop, particularly the functions of memory and of sensory perception, it comes to recognize the absence of its mother (absence of the need-satisfying object) as a prelude or signal that traumatic anxiety may develop. In other words, the infant learns by experience of the danger that when its mother is absent an instinctual need may arise which the infant itself cannot satisfy and which will give rise to anxiety on the basis just described. What happens then is that the danger situation is reacted to with anxiety. This type of anxiety, which is connected with danger or, later, with the anticipation of danger, is called *signal anxiety*, since it



is in effect a signal that a traumatic state may develop. When signal anxiety does develop it sets into operation the pleasure principle, according to which the mind functions in such a way as to avoid pain or discomfort. Thus it becomes necessary that the instinctual need or wish, which might give rise to a traumatic state in the absence of the mother, be prevented from developing into an urgent need. The various ways by which this can be accomplished are referred to in psychoanalytic terminology as the defenses or defense mechanisms of the ego. We shall discuss them in more detail presently.

In normal childhood development there is a regular sequence of danger situations which are associated with signal anxiety. The first of these is the one described above, which is referred to as loss of the object. The situation which next comes to be recognized by the infant as a danger is the possibility of loss of love or affection of the parent on whom it is dependent for instinctual gratification. This is referred to as loss of love. The term "separation anxiety" is often used to indicate anxiety related to either or both of these typical dangers. The third typical danger of childhood, which is characteristic of the phallic or Oedipal phase of instinctual development, is the fantasy of castration in the little boy or of analogous genital injury in the case of the little girl. The term "castration anxiety" designates anxiety related to this danger. The fourth danger, which appears only after the formation of the superego, an event to be described below, has to do with superego disapproval or prohibition. It is sometimes referred to as superego anxiety but more often simply as guilt. In psychoanalytic terminology, the term "guilt" is often used to refer as well to the need for self-punishment and also to the various associated phenomena of remorse, penance, and retribution.

It should be noted that the typical danger situations just described do not simply succeed one another in the sense that the first disappears from prominence in mental life as the second develops. On the contrary, as each new source of anxiety appears in the child's mental life, the earlier ones continue to exist alongside of the new, and all the danger situations are important in everyone, though to varying degrees in each, depending on the life experiences of the individual and possibly on his constitutional endowment as well. [See ANXIETY.]

In considering the defense mechanisms, it is important to recognize that the ego may use anything which is available to it as a defense against an instinctual wish or an unconscious need for self-punishment which arouses anxiety. Thus, for

example, a diversion of attention may serve a defensive purpose, or muscular immobility may be used to defend against the possibility of being sexually or aggressively active. It may even happen that one instinctual derivative is gratified at least partly as a defense against the emergence of another. Thus, homosexual fantasies or even homosexual activity may appear as defense against dangerous heterosexual wishes, or vice versa. In order to avoid possible misunderstanding in this connection, it must be emphasized that, in general, instinctual wishes are reacted to as dangerous in adult life because they were conceived to be so in early childhood. It frequently happens that an adult vigorously defends himself unconsciously against the emergence of a dangerous instinctual wish which, as an adult, he considers either entirely permissible and natural or even trivial and unimportant. Thus it is fairly accurate to say that mental conflict in adult life derives from childhood or, to use psychoanalytic terminology, infantile sources. [See DEFENSE MECHANISMS.]

*Repression.* While bearing in mind that the defensive activities of the ego may be extremely varied, we may mention and define briefly several defense mechanisms which are referred to frequently in psychoanalytic literature by specific names. The one which was first identified and has been most extensively discussed is called repression. In psychoanalytic terminology, repression signifies an active forgetting of the memories and other ideational representations of a dangerous wish, whether instinctual in origin or self-punitive. Repressed memories are no longer accessible to consciousness. Since they remain cathected with psychic energy, however, they are constantly striving for some sort of expression in conscious mental life and behavior, however remote from the original wish that expression may be. Thus a repressed wish or need may exert a considerable, even a decisive, influence in the mental life of an individual. The counterforce which is required to keep such a powerful urge repressed is referred to as a counteractethesis.

*Reaction formation.* Another important defense mechanism is reaction formation. In reaction formation one mental tendency is emphasized in order to prevent the emergence of its opposite. Thus, for example, dangerous aggressive wishes may be kept in check by the development of a character trait of kindness and gentleness. In the same way, the reverse may happen, that is, an individual may be bold and vigorous in his overt behavior partly in order to defend himself against a dangerous tendency to be submissive or passive. Similarly, cleanliness may appear as a reaction

formation against dirtiness, particularly against coprophilia, etc. The most dramatic reaction formations are those in which love appears as a reaction formation against hate, or vice versa.

*Isolation.* The term "isolation," though sometimes used in different senses, generally refers to what is more precisely called isolation of affect. In this defense mechanism, a wish or impulse is allowed to appear directly in consciousness but devoid of emotional significance. It seems to the individual "just an idea" rather than a desire or wish. Isolation, as used in this sense, is closely related to defensive intellectualization in general.

*Identification as a defense.* Identification, a phenomenon of mental life which we have seen to be of general significance in mental development, especially of the ego functions, often plays a defensive role. In psychoanalytic literature much emphasis has been put on the important part played by identification in connection with separation anxiety. However, it also plays a vitally important role in connection with castration anxiety in childhood and is of particular significance in superego formation.

*Projection.* The term "projection" refers to that phenomenon whereby an individual attributes to others wishes or impulses which in fact are his own but which would give rise to anxiety or guilt if he acknowledged them as such. Projection as a defense plays a particularly important part in the psychopathology of paranoid conditions, but it is of considerable significance in normal mental life as well. Prejudices of all sorts—for example, those directed against an enemy in time of war—are often closely dependent on projection.

*Denial and undoing.* The term "denial" is used to refer to a certain defensive attitude toward one or several aspects of the external world. One may fail to perceive an event in one's environment in order to avoid the guilt or anxiety that would result if the event were perceived. Fairy tales and daydreams in general make considerable use of denial. The misperceptions of the environment that play such a large role in many psychoses are mainly related to the defensive use of denial in those conditions.

The term "undoing" refers to actions which are performed in order to undo or render harmless unconscious wishes or fantasies which are considered dangerous. They represent a kind of unconscious magical penance.

*Regression.* Regression is a mental tendency or capacity which often serves a defensive function. The dangers associated with Oedipal wishes or fantasies, for example, may be avoided by regressing to anal or oral wishes. This process is

known as instinctual regression. Regression may also affect various of the ego functions in a defensive way. In general, if regression is permanent and not merely temporary, the greater its degree and the more extended its role in mental life and the more severe the distortion of personality or the mental illness that results.

### The Oedipal phase and superego formation

At the age of two and a half or three the child normally enters into a phase of psychosexual development called the Oedipal period or phase. This lasts for about two and a half or three years and corresponds approximately with what we have earlier referred to as the phallic phase of libidinal development. The violent intrapsychic conflicts which are caused by the sexual wishes of the Oedipal period are of crucial importance in mental development and later mental functioning. They profoundly influence character development in general and are responsible for superego development and organization in particular. In addition, they form the basis or groundwork for the vast majority of whatever neurotic symptoms may appear in later life. [See DEVELOPMENTAL PSYCHOLOGY; PERSONALITY, article on PERSONALITY DEVELOPMENT.]

Freud used the term "Oedipal" to designate this period of mental development because of the similarity between certain of the wishes that characterize it in the little boy and the legend of Oedipus as recounted in Sophocles' play. The little boy in the Oedipal period normally is passionately enamored of his mother, wishes to marry her, and to displace his father in his mother's bed. The corresponding wish to marry her father and take her mother's place is equally characteristic of the little girl in the Oedipal period. Thus, the sexual wishes of the Oedipal period revolve around the fateful themes of incest and parricide.

In fact, however, the wishes and conflicts of the Oedipal period are somewhat more complex than the name would indicate. In his psychosexual life, man is clearly bisexual. This is nowhere more evident in the normal human being than during the years of early childhood. In the usual family constellation the little boy of the Oedipal period has sexual feelings toward both parents and jealously murderous wishes toward each, although normally the heterosexual wishes and the jealousy connected with them considerably outweigh the homosexual ones. The same is true for the little girl during her Oedipal period.

For a variety of reasons, not all of which are clearly understood, the Oedipal boy fears castration, specifically the loss of his penis, in connection



with his incestuous and parricidal wishes, and the little girl during the same period fears some analogous genital injury. In the case of the girl, the conflict of feelings is complicated by an awareness that she has no penis, a fact which she regularly interprets as a sign of inferiority. Because of the castration anxiety aroused by their instinctual wishes, children of both sexes attempt to control their Oedipal wishes with a great variety of defensive reactions. [See SEXUAL BEHAVIOR, article on SEXUAL DEVIATION: PSYCHOLOGICAL ASPECTS.]

Among the most important of these for the future development and functioning of the child's mind are the reactions which give rise to what is called the superego. This group of mental functions may be defined as comprising those functions of the mind which are concerned with either moral prohibitions or exhortations. The functions which have to do with exhortation are sometimes referred to as the ego ideal. Activity of superego functions results either in a limitation of behavior to morally acceptable modes of instinctual gratification or to that particular state of unpleasant inner tension which is variously referred to as guilt or remorse and which leads to acts of penitence, of reparation, of self-punishment, and—in the most extreme instances—to self-injury or self-destruction.

**Identification and the superego.** The principal mental mechanism involved in superego formation is identification. Normally, the little child identifies with the prohibiting and punishing aspects of the parent who is his chief rival. By making the latter's threats and prohibitions his own, he defends himself against the fantasied danger of castration. He thus achieves a considerable degree of control over the instinctual wishes which for him constitute an overwhelming danger. At the same time, however, this identification results in an appreciable degree of permanent limitation of opportunities for instinctual gratification. Thus, superego formation normally results in a considerable degree of interference with the operation of the pleasure principle, or at the very least it considerably complicates its operation.

It follows from what has been said that the core of every individual's morality and, consequently, the invariant elements that one would expect to find at the center of the moral code of every human society, consist of a prohibition against incest, particularly between parent and child, and against parricide. One may add that for many, or perhaps most, individuals, one of the results of superego formation is a considerable degree of guilt over genital masturbation, since this is usually the physically gratifying activity which accompanies

the sexually exciting wishes of the Oedipal period. It is also interesting to note that the threats and punishment associated with superego functioning in early childhood, and persisting unconsciously into later life, follow the principle of *lex talionis*. It would appear that the mind of the child, like the mind of the primitive, follows the principle of an eye for an eye and a tooth for a tooth.

It seems correct to say that the major share of superego formation is the result of the fateful conflicts of the Oedipal period. It should be noted, however, that a considerable degree of superego formation occurs gradually during the ensuing years of childhood. Some of the identifications with teachers and with childhood heroes, whether historic or contemporary, contribute their share to the growing child's superego. It is also true that even in adult life, substantial changes may take place in superego functioning, sometimes in a dramatic way, as for example in cases of religious conversion. Freud also emphasized the changes in superego functioning that are characteristic of mob psychology. According to Freud, identification with a group leader—for example, with a political dictator—may produce as considerable an alteration in the superego functioning of many of his followers as results from religious conversion. [See MASS PHENOMENA; MORAL DEVELOPMENT.]

**Latency and adolescence.** The institution and organization of the superego normally result in a considerable subsidence of instinctual conflict beginning at about six years of age. Although this subsidence is by no means complete, still the whole picture of a relative quiescence of psychosexual activity and conflict during the years between six and twelve is in considerable contrast to the turbulence and conflict so characteristic of the Oedipal period, which precedes it, and of adolescence, which follows it. Freud proposed to call this interim period a period of latency. When the physical, particularly the sexual, changes of adolescence begin, there is a recrudescence of sexual wishes and sexual conflicts. The course of these adolescent stirrings and conflicts is to a considerable degree predetermined by the defenses already established during the Oedipal period. Thus, to a greater degree than is usually appreciated, the psychosexual development of adolescence and early adult life is a recapitulation of childhood psychosexual development. [See ADOLESCENCE.]

The foregoing is an extremely brief summary of psychoanalysis as a general science of behavior. Because of its brevity, it is inevitable that the reader may often be misled in attempting to under-

stand a subject which is at once so complex and so extensive. This is not, however, the most serious source of difficulty to the reader. A much greater difficulty is likely to arise from two rather different sources. The first of these is a lack of familiarity and experience with the use of the psychoanalytic method. As mentioned in the introduction, the data concerning human behavior which are available only by application of the psychoanalytic method are, generally speaking, quite foreign to the experience of anyone who has not had an opportunity to apply it. Lack of such opportunity, therefore, makes it difficult for one to form an adequate judgment concerning the validity of many psychoanalytic propositions. The second source of difficulty results from what astronomers call the personal equation. In the case of childhood sexuality, the personal equation is very great indeed. Each adult individual, as a result of having passed through the conflicts of the Oedipal period himself, has stubborn defenses within himself against his own childhood wishes. Even to recognize their existence in others is likely to arouse considerable guilt and anxiety. It is much more comfortable—that is, less frightening—to minimize the importance of childhood sexuality in human mental life or to deny its significance altogether. It is no accident, as Freud pointed out, that every adult has an extensive amnesia for the events of early childhood. Most individuals can overcome their inner conflicts over childhood sexual wishes only with the help of personal psychoanalytic treatment. For this reason, a personal analysis is a usual prelude to psychoanalytic practice, that is, to the systematic, usually therapeutic, application of the psychoanalytic method.

The unavoidable limitations which, until now, have been imposed by the above facts on the acceptance of psychoanalysis by many social and behavioral scientists are unfortunate. The fact is that psychoanalysis can supply to the social and behavioral scientist, as to the student of art and literature, a more profound, more complex, and more accurate knowledge of the nature of man—his needs, his fears, his conflicts, and his motives—than is available from any other source. Such knowledge is frequently important and often vital. It should properly be part of the education of all who deal professionally with man and his works.

CHARLES BRENNER

#### BIBLIOGRAPHY

ARLOW, JACOB A.; and BRENNER, CHARLES 1964 *Psychoanalytic Concepts and the Structural Theory*. New York: International Universities Press.

BRENNER, CHARLES 1955 *An Elementary Textbook of Psychoanalysis*. New York: International Universities Press.

FENICHEL, OTTO 1945 *The Psychoanalytic Theory of Neurosis*. New York: Norton.

FREUD, SIGMUND (1895–1926) 1953–1963 *The Standard Edition of the Complete Psychological Works of Sigmund Freud*. 24 vols. London: Hogarth; New York: Macmillan.

HARTMANN, HEINZ (1939) 1958 *Ego Psychology and the Problem of Adaptation*. Translated by David Rapaport. New York: International Universities Press. → First published as *Ich-Psychologie und Anpassungsproblem*.

HARTMANN, HEINZ 1960 *Psychoanalysis and Moral Values*. New York: International Universities Press. → Freud anniversary lecture, given in 1959 at the New York Psychoanalytic Institute.

KRIS, ERNST 1947 *The Nature of Psychoanalytic Propositions and Their Validation*. Pages 239–259 in Sidney Hook and Milton R. Konvitz (editors), *Freedom and Experience*. A New School for Social Research publication. Ithaca, N.Y.: Cornell Univ. Press.

RUITENBEEK, HENDRIK M. (editor) 1962 *Psychoanalysis and Social Science*. New York: Dutton. → A useful selection of essays on the impact of psychoanalytic thought upon anthropology, sociology, and the social sciences generally.

WAEELDER, ROBERT 1960 *Basic Theory of Psychoanalysis*. New York: International Universities Press.

#### II

#### EGO PSYCHOLOGY

Although the term “ego” has meanings that go back to the beginnings of philosophy, it is only its place in psychoanalytic theory that will be the concern of this article. For this reason it is necessary from the start to rid the term of some of its popularly understood, nonpsychoanalytic meanings. The ego is not simply intelligence; neither is it an inflated sense of one’s own importance, nor does it refer to its traditional philosophical meaning of self-consciousness. It was not meant by Freud to be strictly a phenomenological concept. The ego is not simply the subject of experience, in contrast to the objects of experience, nor is it one’s own person (self), in contrast to others. These exclusions are neither arbitrary nor a matter of preference, but reflect the place that the concept is meant to fill in psychoanalytic theory: specifically, Freud’s need for suiting the meaning of the term “ego” as a “representative of reality” to the requirements of the parent body of psychoanalytic theory. Thus, in Freud’s last and most fully developed conception of the ego, self-feelings, the capacity to reflect on oneself as a person, and the capacity to experience distinctions between self and others are *functions* within the ego system, part of the ego’s adaptational repertoire of controls and defenses as a subsystem of the personality. The ego’s fundamental mean-



ing for Freud was a hypothetical array of processes standing generally for self-preservation and survival and having to do with the structuring of purposive action in relation to various aspects of reality.

This is not to say that psychoanalytic theorists are always clear and consistent in their usage. Sometimes, the term is merely a taxonomical device to refer to defensive and adaptational motives and their instrumentalities. Sometimes, it is used in the more ambitious connotation of a distinctive subsystem of the personality whose regulatory functions extend to the control of two other subsystems, superego and id. In this sense it becomes more an independent variable that has a role in "affecting" behavior; it is not simply a taxonomy of motivational phenomena or of functions but an entity whose parameters, mode of operations, and functions are potentially capable of being specified in some model of self-regulation.

Confusion in psychoanalytic usage developed from the fact that while the ego concept was present from the very beginnings of Freud's theory, it underwent a number of changes and shifts of emphasis in response to developments in psychoanalytic theory at large, a fact which Freud sometimes failed to acknowledge. Part of the intent of this review is to draw attention to invariant meanings that have clung to the term throughout its history in psychoanalysis, as well as to changes of emphasis. Of equal concern will be the logical place of the ego concept within psychoanalytic theory and its distinctive character as a framework for explaining behavior. Since the history of the ego concept has been described several times (e.g., Hartmann 1939; Kris 1951a), we will touch on historical considerations only to the extent of indicating how changes in the ego concept reflected conditions in the parent body of psychoanalytic theory—particularly the gradual emergence of the ego concept from its setting within a theory of illness to one concerned with normal development—and to the present status of the concept as constituting a distinctive field of psychoanalytic enterprise called ego psychology. A final and important consideration will be an assessment of its current status in psychoanalysis.

### The ego in Freud's earliest model

From the start, Freud needed a concept of ego. Indeed, it is a curious fact that his concept antedated his theory of psychosexual drives, generally considered to be the cornerstone of psychoanalysis. It is fair to say that Freud's early theory of the mind was as unavoidably dependent on a concept

of ego as his later theories were on a concept of instinctual drives. Believing that he had discovered the primary etiological basis of hysterical symptomatology—a psychical "lesion" traceable to a traumatic sexual experience in childhood—he envisioned, in a burst of temporary enthusiasm, the mechanics of neurosis, within a comprehensive neuropsychological model of normal behavior. In a work published posthumously as *Project for a Scientific Psychology* (1895), he developed his first systematic account of the fundamental forces in man's behavior. Behavior, he said, is ultimately the expression both of the discharge of pressurized build-ups of internally and externally generated tensions through appropriate and effective actions upon the environment and of structural complications that develop in the organism in the course of its efforts to find the most effective means of discharge. He called these the primary and secondary functions of psychical structures, respectively.

The ego was conceived as an essential structural means of accomplishing the secondary function—the agency of effective discharge, one that could respond simultaneously to the demands of energetic tension and to environmental conditions suited to tension reduction, that could monitor motoric instrumentalities of discharge (because effective discharge is achievable only through action) and that could be implemented by perceptual-cognitive operations of thought which guarantee the proving of reality or "reality testing." The ego, then, was a coherent organization that was endowed with capacities for inhibiting energy flow and that operated in coordination with processes having to do with the "representation of realities"—thought, perception, judgment, reflection, etc. Freud even specified a latticelike arrangement of neurons with special capabilities of retaining excitation and directing energetic transmission. In the *Project*, defense, which later was to be a dominant property of the ego, was construed, in both a broad and a narrow sense, as part of the secondary function: in the broad sense, as an inhibiting, but adaptational, process which directs the course of discharge and, in the narrow sense, as a specific form of inhibition or "repression." The point is of interest because for two whole decades following the *Project*, the latter, narrow sense of "defense" preempted all other meanings of "ego." In this remarkable document, which merits more detailed study than is possible here, Freud spelled out the functions of the ego, which he was not again inclined to do in as systematic a way for some thirty years. [See DEFENSE MECHANISMS.]

### Status of the concept: 1900–1923

In what is generally regarded as the second and main period of Freud's psychoanalytic writings—the period immediately following his writing of the *Project*, up to around 1923, and including *The Interpretation of Dreams* (1900)—Freud's theoretical interest in the ego concept waned, although it frequently appeared as a necessary descriptive term. Functions that had been specifically assigned to the ego in the *Project* were now dealt with separately, without being explicitly associated with the ego. The term had a variety of descriptive usages, sometimes with the connotation “conscious inhibitive control,” sometimes with the meaning “motivations concerned with self-preservation” (e.g., “ego drives”), sometimes in the sense of “the internalized ideals and values of identification figures,” and sometimes in the sense of “feelings of self.” Mainly, however, it seemed to have the narrow connotation “an instrumentality of defense” and specifically of repression. In this last sense, the ego would refer variously to the conscious acceptable mass of ideas which took part in the conflict, the ideals which precipitated the conflict, or the resisting force in conflict, the agency of the regressive process itself. From the vantage point of both his earlier and later, final conception, these were all *partial* attributes of the ego concept. The idea of a comprehensive, coherent system of control was put aside.

Several reasons have been put forward for this and for the generally unsystematic way in which Freud used the concept in this period. In part, this reflects Freud's style of theorizing. He tended to give main theoretical weight to his immediate empirical concerns, often without attention to the resultant, sometimes dissonant juxtapositions to earlier formulations. The lure of the next bend, beyond which new discoveries might be made, was always distracting him from the labor of theoretical inventory. It is, of course, also true that he wanted to rid himself of the neuronal model, which he considered a failure, and with it went his early strategy of trying to encompass pathological phenomena within a more comprehensive model of *normal* functioning. Instead, he now gave himself over entirely to the observations and phenomena that he was most immediately confronted with—those of illness, symptoms, and conflict. He not only moved toward a physiologically “neutral” form of theory but also became more specifically and dominantly preoccupied with the dynamics of *illness* and less with the task of a comprehensive *general* psychology of adaptation. Concern with

illness meant a theory of its genesis in conflict and focusing on the roles of sexuality and instinctual drives. In the *Project*, he had tried to view issues of discharge and control, not simply in relation to sexual wishes and tensions, but in the broader perspective of a general biological enterprise of adaptation to stimulation. There the secondary function of defense was broadly construed; now it became identified with the ego in the narrower sense of repression produced by conflicts in which sexual wishes and desires were pre-eminent. The mental lesions—obsessions, hysterical symptoms, perversions, and the like—and their treatment were placed in the forefront, and elaboration of the principles of neuroses, conflict, and the relation of both to sexual bases became the core of Freud's theoretical enterprise. In this connection two developments were of monumental importance in setting the direction of his work for the next two decades: his retraction of the etiological importance he had given to sexual trauma in childhood and his altered views of the nature of the conflictual bases of neurosis.

His first conception of the importance of sexuality in the causation of hysteria (1896a) held that the memory of an actual traumatic sexual experience in childhood (e.g., seduction) was central; in this view, the memory of such a prepubertal experience becomes especially pathogenic in the later period of genital sexuality, creating the condition of neurosis-inducing conflict. However, he gave up the conception that the critical etiological event was an actual seduction when he became convinced that the disturbing idea in hysterical neuroses was not a memory, but a *fantasy* of seduction woven out of a tabooed sexual wish.

This had repercussions on the very fundamentals of psychoanalytic theory, for now it became important to ask how such a fantasy and the wishes involved in it came about. The three main sources of neurotic illness—the sexual core of conflict, the unconscious wishes, and the unsuccessful repressive defense against these wishes—leading to symptoms henceforth became the main themes of Freud's theoretical efforts. He embarked on observations of continuities in sexual experiences at different stages of development and of their pathological miscarriages. His clinical data pointed to the autogenetic connection between both early infantile needs and bodily functions and the later development of genital gratification.

It was in this connection that the theory of instinctual sexual drives became a critical anchorage of Freud's thought—the all-determining “reality” to which adjustment must be geared; an internal



stimulus which, being inescapable through flight and being peremptory and periodic, creates a continuing and unrelenting source of wishes and strivings that has to be discharged through appropriate "objects." The other vital appetitive instincts were not very important, because they seemed to contribute little to an understanding of neurosis. By far the most important source of motivation and of individual development was the sexual drives and the conflicts engendered by their momentum for gratification. Moreover, the sexual drives underwent a developmental change in respect to their aims (wishes) and their potentialities for possible gratification (objects). Indeed, they provided the primary momentum of growth. So great was their structuring power that Freud tended at times to believe that character development itself is patterned exclusively by fixation upon different aims and modes of drive gratification. The environment was accorded little significance as a participant in the development of instinctual drives—only in the conflicts generated by them; family life and rearing create obstacles to the discharge of drives, thereby necessitating indirect forms of discharge or total repression.

For Freud, by far the most crucial discovery was that of infantile sexuality—that is, that there is a developmental continuity in requirements of sensual gratification from infantile modes to the usual adult forms of sexuality. Freud wrestled with the formidable theoretical problem of how to fit into a uniform conception the facts of an erotic momentum, its seeming continuity in development, interchangeability of sensual aims and modes of gratification, and pathological deflections into perversion, fixations, and regressions. Freud's theoretical solution was to conceive of an active principle of sensual development, involving an energetic quantum, which he called libido, having distinctive qualitative or directional properties. There is inherent in libido an implacable developmental course in respect to the aims of its discharge, these developmental changes appearing in stagelike progression. At the same time, within each stage, libido is capable of some degree of transformation of aim and of displacements in respect to actual objects of discharge. For example, "conscientiousness" could be an instance of the transformation of "anal libido." Libidinal discharge became for Freud the most important single basis of motivation. He believed that the requirements for the relief of libidinal tensions, the conflicts to which they give rise, the shifts in distributions of libido, and the deflections in their discharge are not only the main

issues in neurosis but the essential paradigm for picturing normal motivational development.

Through the simplification achieved by attributing directional attributes to a distinctive type of energy, Freud had an attractive and seemingly efficient means of coordinating the complex facts of sexual development and the autogenic heightening and lessening of the desire for sensual gratification in each period. Through the *quantitative* attributes of libido, the theory could account for intensity of drives, for the pressureful momentum of sensuality; through its *qualitative* attributes of directionality, it was possible to account for selectivity in respect to the objects which can elicit its discharge. Libidinal energy in its various drive forms had characteristics usually associated with animal instinct; at the same time, it had the crucial property of displacability, which distinguished it as specifically human. Thus, the forms of transformation from oral to anal, to phallic, and to genital stages are inherited, like the momentum of transition from one phase to the next, but within each stage the options of gratification are considerable. It must be noted, however, that the distinction between the energy, or activation, aspect of sensuality and the executive actions associated with its formation into drives was obscure in Freud's formulation—and has remained so. This uneasy coexistence of quantitative and directional assumptions in the libido theory and their unclarified relationship with each other remain weak points in psychoanalytic theorizing to this day. It should be noted that Freud was also veering away sharply from his early emphasis on the importance of environmental (exteroceptive) stimulation.

Sharing the spotlight with libidinal drives at this time was the nature of conflict in neurotic disturbance. For Freud a key factor in neurotic conflict was repression—the blocking of a wish and its derivatives and associations from conscious understanding and action; symptoms of conflict were a manifestation and outcome of repression—they reflected a strategy of conflict resolution arising out of the impasse which repression imposes upon gratification. That his interest narrowed to the one adaptational function of defensive repression partly reflected the logical momentum of his theory which led to, and was continually fed by, particular emphases of clinical observation. It is also true, however, that clarification of the dynamics of repression seemed to be dictated by the therapeutic necessity. For according to the theory, the goal of therapy is to uncover the repressed; therefore, it is crucial to give observational priority to indications of the

"return of the repressed" and to the means of overcoming the ill effects of repression. [See CONFLICT, article on PSYCHOLOGICAL ASPECTS.]

Although the ego was present in this conception, its status was that of a curiously vague entity, identified most of the time, but not exclusively, as the agency of the repressive force. Emphasis was given to its *resistive* function, in relation to drive and wish, and its *inhibitive* capacity, rather than to its capacities for promoting effective gratification; it was identified not only with the repressing force but with the body of ideas—internalized value structures, as well as consciously acceptable and accessible ideas, which instigated resistiveness to drive discharge. Freud also tended to link capacities for conscious experience and will with the ego, but this was loosely developed and does not easily fit with the notion of the ego as an agency that promotes elisions of consciousness through repression. The absence of any detailed regard for how these oppositely effective functions are related within the ego is an example of Freud's disinclination, in this period, to systematize the ego conception.

It is possible that the reduced importance of a specific environmental etiology of neurosis affected Freud's interest in the ego, if not his assumption that the ego was necessary in his over-all conception of psychical structure; that aspect of the ego which is purely adaptational became less compelling. Parental behavior and family life were significant aspects of the environment—but significant only as potential impediments which produced conflict and provoked defense. The possible supportive importance of environmental structure to normal drive functioning as a molding and implementing influence—not simply an obstacle to the evolutionary-given momentum of libido—was not a natural fulcrum of interest for libido theory and, hence, received little attention. One also wonders if premonitions of possible contradiction with the libido theory contributed to the stasis of the ego concept in this period, since Freud's account of the motivational and structuring force of libidinal drives included attributes that he later, under the pressure of observation, imputed to the ego.

As for motivations specifically attributable to the ego, Freud spoke about them only with respect to the ego's role as an instrumentality of the pleasure principle and its role of defense. He did for a time speak of "ego drives," but he gave up the notion, possibly because of difficulties in making it compatible with the libido theory. It seemed necessary to emphasize only *reactivity* to instinctual drive and

containment of drive by defense. Indeed, this emphasis is common even today. Clinicians are still inclined to regard the ego mainly in the light of its participation in conflict and in its partnership with instinctual drives pressing for discharge.

Freud, nonetheless, had much to say about thinking processes in their adaptive, as well as drive-related, aspects. However, he did not bring these considerations into a single taxonomic concept of an ego system of control. Thus, in *The Interpretation of Dreams* appears his epoch-making distinction between primary and secondary thought processes. Secondary thought shows characteristics that follow the precepts and rules of logic, maintaining an identity or coordination with reality structures and events; in marked contrast, thought processes of the primary process are wholly in the service of drives, of promoting peremptory discharge. The distinction implies that there is an "objective reality" and that thought (in the secondary process) is capable of representing it without distorting it. However, Freud left largely unexplained and undescribed the developmental origins and nature of the secondary process. In fact, it never again received the attention he had given it in the *Project*.

It is possible that some of the difficulties that have persistently clung to the ego concept since Freud's time are traceable to the fact that the secondary process and its reality-testing functions were not explicitly made a part of ego theory in this period; its status as a means of implementing effective drive control, as well as a means of representing reality relationships generally, was implicit but not developed. Also, the question of whether the mechanisms of symbolization, condensation, and displacement of the primary process are best viewed as ego manifestations to this day is unresolved. To have confronted such questions—and generally that of making primary and secondary processes comprehensible within a conception of ego—would have drawn attention to the crucial issue of the independence of ego processes from libidinal control. Moreover, there was something contradictory in the view of the ego as primarily a reactive, defensive agency whose vital biological mission is to insure self-preservation. There are, in his writing, any number of hints that it was difficult to avoid the assumption of an autonomously active force in the ego. The critical point is that Freud did not consider in detail the problem of instrumentalities of control in general, whether in the service of drive or in nondrive adaptations. Had he done so, we would have had a psychological conception more consistent with the then cur-



rent, as well as present, preoccupations of most psychologists.

It seems fair to conclude that during this period the assumptions of his libido theory muted Freud's interest in an interaction of adaptive motivations and environment as a major motivational source of behavior. It required repeated clinical indications of the explanatory insufficiency of the theory of libidinal drives for him to widen the orbit of the ego's meaning to embrace these phenomena.

#### Later development: 1924-1938

The decision to confront more directly and systematically the ego's position in psychoanalytic theory was gradually forced upon Freud. Actually, this renewed interest picked up themes of the early *Project*. Clinical observations and considerations had much to do with forcing the issue—facts and special theoretical formulations that had no firm position in the general theory as it stood had been accumulating. We will consider several of these: (1) phenomena of unconsciously manifested resistance in psychoanalytic treatment; (2) a radically different appreciation of the meaning of anxiety; (3) the role of unconscious guilt and internalized parental commands and morality; and (4) psychotic breakdowns of thought functions.

**Unconscious resistance.** Freud became increasingly impressed with a peculiar resistance to being helped or understood that patients displayed. The patient could proclaim with conviction his willingness to participate in the treatment, but there were many indications that he was undermining it in a variety of ways. The patient seemed to resist knowing not only the repressed ideas but, more importantly, resisted knowing and acknowledging that he was repressing. The resistance seemed to operate quite unconsciously, indeed seemed even to gain in force from the fact that it was not conscious. These facts profoundly affected the strategy of therapy itself and also moved Freud toward a reformulation of the ego concept.

Accordingly, a great change eventually came about in psychoanalytic therapy—an emphasis upon working with defenses themselves and with resistances; a great deal of attention was now given to their various forms and vicissitudes in treatment. No longer was analysis considered to be simply the uncovering of repressed memories through catharsis. The conception of repression itself underwent change. Repression is a process that not only affects memory but promotes breaks in the conscious appreciation of associative links with the conflictual core of the neurotic symptoms; such "splits" produced by repression have different be-

havioral manifestations. Thus, even the meaning of uncovering repression became different; it could mean not only recovering a memory but the re-establishing of a conscious appreciation of causal connections among ideas.

The phenomena of resistance seemed also to point to an active principle in the structuring of behavior, independent of, and rivaling, that of instinctual drive. Defense could not simply be a reactive response fashioned out of repeated frustrations of drive discharge and modeled upon the discharge modes of instinctual drives. Moreover, the unconscious aspect of this seemingly drive-resistive force cast doubt upon the logic of Freud's previous tendency to associate all unconscious determinants with instinctual drives alone.

**The new theory of anxiety.** Freud revised his views of anxiety, and his new conception had reverberations throughout the body of psychoanalytic theory and therapeutic practice. Previously, Freud had resolved the question of whether anxiety is a reaction to danger or a transformation of libido, by what seemed a parsimonious solution in terms of libido theory alone. Anxiety was simply a product of libidinal drive blocked from discharge, an overflow of libidinal energy which induced a passive reaction of inchoate restlessness and diffuse feelings of fearfulness. However, a number of clinical considerations—among them, the facts of resistance and the problem of how repression itself is initiated—compelled him to assign an active role to anxiety in a regulative system of response to danger. He proposed that small quanta of anxiety had utility as a signal, an informational sign of impending danger, in response to which defenses are alerted. Thus, anxiety became associated with a directive function distinct and separable from libido and able to determine the course of behavior independently of libido. He underscored the separation of libidinal drive discharge, but he avoided confronting the issue (If it did not originate from libido, whence the energetic basis of anxiety?) by making it an ego function and part of the ego's regulative system of defense.

The importance of this new view of anxiety can hardly be overemphasized; it brought into the foreground an anticipatory capacity in the schema of regulation. Effective regulation meant functioning guided by anticipated danger. The anxiety signal is more than a realistic response of fearfulness to actual danger. It reflects a regulatory system capable of initiating defensive action on the basis of a premonition; it indicates a system capable of the developmental feat of turning passively experienced, full-blown anxiety into a finely tuned means

of anticipatory defense. Moreover, the anxiety signal is capable of mobilizing a great variety of defenses in this fashion.

The new theory also emphasized the qualitative, or configurational, aspects of anxiety, in contrast with the earlier, purely quantitative energetic conception. In general, anxiety is indicative of a state of mind in which a person is preoccupied with preserving, rather than with gratifying, himself. Anxiety is linked to internal representations of danger situations; these prototypic dangers differ at the various stages of psychosexual development, producing different anxiety configurations. For example, separation can provoke in the infant an experience of loss of love, which is internalized as a potential danger and is thereafter an occasion for anxiety. Separation and object loss are especially potent sources of anxiety in the period of infant helplessness; later, the dangers of retribution, castration, and fantasied loss of potency become sources of anxiety. Freud, thus, implied that the controlling system which embodied the anxiety signal had a developmental aspect coordinate with, and separate from, that of libidinal development, one in which different focuses of anxiety at different ages result in correspondingly different defensive patterns. The theory also led to a further elaboration of neurosis as the retention and reactivation of past or obsolete conditions of anxiety—for example, loss of love, castration, etc. [See ANXIETY.]

The importance now attached to developmental circumstances that promote anxiety broadened psychoanalytic concerns to include the bases and content of ontological insecurities, their epigenesis and relation to social development. It began to turn attention to the patient's personal construction of reality—that is, reality as he interprets it, whether of environment or of drive. The impact on psychoanalytic therapy was great and helped to redirect it. Now, in addition to asking what the patient desires, one has to ask what he is afraid of and what he does if he is afraid. Only by understanding the configuration of his anxiety and the modes of control associated with it can one get a complete picture of his conflicts and fears.

But the new conception posed a difficult problem for existing instinctual-drive theory. Since the link between anxiety and libido theory has been weakened by attributing anxiety's essential quantitative factor only vaguely, if at all, to libido, how does one relate anxiety to the parent theory? Freud resolved the issue, not by revision, but by addition: by broadening the ego conception to include properties of activity that he had not before assumed

it to have and by including anxiety among its signaling capacities. He thought that by endowing the ego with its own energy, he had an expedient solution which did not threaten the existing edifice of psychosexual theory, a solution which seemed to retain its explanatory usefulness and to accommodate certain observations that had stubbornly resisted interpretation.

**Unconscious moral strictures and guilt.** Observations of patients who hold on to suffering out of a need for punishment and of the phenomenon of guilt seemed also to be better provided for if one assumed some grade of regulative capacity independent of instinctual libidinal drive. The fact, too, that the qualms of conscience can be effective without being conscious called further attention to the inadequacy of a theoretical bias Freud had shown toward equating conscious and unconscious events with the polar participants in conflict. He now assigned motivational importance to this facet of regulation and termed it superego. The superego gives commands; it is an internal "must" system which can mount urgent motivational pressure. However, after so internalizing and elevating the motivating forces of a morality system, Freud had to cope with conceiving another system that was responsive to it, one that was responsible for modulating and implementing its demands. Freud called upon the ego concept to provide the supplement: the activity of internalized moral strictures and values became part of a general system of ego control, which is coordinate with the id and which is capable of imposing restraints, of staking out the range of permissible drive discharge, of determining different forms of conflict, and of bringing about their resolution. [See MORAL DEVELOPMENT.]

**Psychotic thought disorders.** Psychoanalytic experiences with psychotics pointed to impairments of adaptive capacities and to defective personal constructions of reality as primary issues. For example, the difficulties that psychotics have in sustaining emotional investments (transference) seemed to Freud to indicate a fundamental adaptational deficit in the affective sphere, which was itself a precondition, rather than a result, of the breakthrough of drives into cognitive forms (the primary thought processes). Clearly, clinical observations increasingly pointed up limitations of the existing model, which regarded libidinal drives as the dominant regulatory force in behavior, counterpoised by a vaguely defined inhibiting ego.

### Expansion of the ego concept

In confronting the insufficiencies of the general theory, Freud could have revamped his conceptions



of psychosexual (libidinal) determinism, conflict, and defense. The solution Freud adopted consisted, instead, of broadening the ego concept; he converted it into a hypothetical system of regulation endowed with an activity principle of its own, one that was coordinate with instinctual drives. All processes of selectivity of thought and action in their adaptive aspect were assigned to it. No longer was the ego simply an *ad hoc* passive outcome of impulses meeting implacable external barriers to discharge. This seemed to take care of the principal requirements of the new data—providing for adaptational motive tendencies and for decision, choice, and integration; control of drives was now viewed as but one aspect of the organism's development. Moreover, the new schema brought to psychoanalytic theory a symmetry that it had previously lacked, for now the two main regulative principles, the pleasure and reality principles, were unequivocally associated with two coordinate and corresponding structuring instrumentalities—the pleasure principle with the id and the reality principle with the ego (and superego) system. Freud stated his final conception of the ego as follows:

Here are the principal characteristics of the ego. In consequence of the pre-established connection between sense perception and muscular action, the ego has voluntary movement at its command. It has the task of self-preservation. As regards *external* events, it performs that task by becoming aware of stimuli, by storing up experiences about them (in the memory), by avoiding excessively strong stimuli (through flight), by dealing with moderate stimuli (through adaptation), and finally by learning to bring about expedient changes in the external world to its own advantage (through activity). As regards *internal* events, in relation to the id, it performs that task by gaining control over the demands of the instincts, by deciding whether they are to be allowed satisfaction, by postponing that satisfaction to times and circumstances favorable in the external world or by suppressing their excitations entirely. (*An Outline of Psychoanalysis* [1940] 1964, vol. 23, pp. 145–146)

Through the critical functions of inhibition, detour, postponement, and integration via thought and action, the functions of defense, of realistic drive implementation, and of adaptation were joined to a single regulative system, coordinate with the drive system itself. Perhaps the most critical capacity attributed to the ego was an autonomous integrative function (actually developed more explicitly by Hartmann) that specifically cast the ego in a much more active role than Freud had ever before accorded it. According to Hartmann, "The most incisive change which took place in

Freud's model of psychic personality . . . [was] its representation as a (more or less) integrated whole, subdivisible in centers of mental functioning—these substructures being defined by their functions, and their demarcation being based on the fact that empirically he found greater coherence among some functions than among others" (Hartmann [1927–1959] 1964, pp. 146–147). This reference to "centers of mental functioning" represents the "structural point of view" in psychoanalysis.

Adjustment, control, and integration, and not simply the direct and indirect forms of drive discharge, thus became the main concerns of psychoanalytic theory. The achieving of harmony, balance, and general equilibrium, as well as the equilibria which result from conflict-determined defensive maneuvers, became a legitimate and even necessary focus of psychoanalytic observation. There is even a hint that the ego not only is coordinate with drives but also serves as organizer of drives, as a kind of omnibus regulator in the personality.

In general, the key change in Freud's conception of the ego from his previous uses of the term was his recognition of an active principle independent of libidinal drive energy, which involves the turning of *enforced* responses into *ego-initiated* processes. The new conception brought into the foreground the proposition that the causal matrix of human behavior cannot be limited to unconsciously active instinctual drives and was the first explicit recognition of this in Freud's thinking. The motivations that develop from this activity are "ego motivations" as distinguished from "drive motivations." The ego's dominion over the pleasure principle became, as Hartmann remarks, a possibility that was unthinkable in the earlier theory.

Indeed, so far did Freud go in his appreciation of the regulative power of the ego that in a later, sobering reflection on negative therapeutic reactions in analysis, he even proposed that there might enter into a person's ego development a peculiar characteristic or deformity of its growth that predisposes it to conflict, whatever the drive components of the conflict—an inherent tendency of the ego to promote the conditions of conflict. Clearly, it was implied that the ego has a degree of organizational independence giving it capabilities of motivating behavior independently of drives.

Although Freud intended to solve problems in the general theory, it is ironic that his respect for the organizing momentum of ego activity actually created a dilemma in psychoanalysis, raising the critical problem of what the energetic basis of the

ego's autonomy is. Freud's hint that the ego was energized by a "neutral," nondrive quantity had the effect of sharpening issues of incompatibility with libido theory, which his expanded ego conception was intended to resolve. It has raised questions about the tenability of his drive and energetic conceptions as explanatory tools. We will return to these issues.

Freud's final conception of ego created the prospect of a new field of psychoanalytic endeavor concerned with ego development; it drew attention to a primary biological fact: a protracted period of helplessness in the child, which had a great influence on the development of adaptational apparatuses and controls. Psychoanalysts were urged to consider in more detail the gradual emergence of control and internalized values that are attached to relationships and the manner in which these become means of regulation. The purely ego aspect of "object relations" thus became an area of possible study.

Freud's final conception, although clear enough in its outlines, was hardly more than a series of programmatic hints and was ambiguous in a number of respects, such as the following. (1) The line of distinction between conflict and adaptational problem was by no means clear. The ego, in the new conception, was seen as an active adaptational participant, implying that neither all encounters nor all experienced impasses necessarily involve conflict. But, the distinctions between disequilibrium and conflict—and, for that matter, between the varieties of disequilibria and of conflict themselves—were by no means clear. (2) Although the new conception implied a distinction between specifically defensive encounters and other modes of adaptational controls, Freud still tended to emphasize the ego's regulation of conflict through defense, although it was an ego working with "its own energies." However, the new conception encouraged the broader view that behavior has a "multiple function"—that the organism is responsive to a complex array of "realities," not simply to the drive-defense polarities of conflict. But Freud had said little about the varieties of adaptational encounter or about developmental "tasks" other than those involving instinctual drives. (3) There was still no explicit recognition of self as a grade of ego organization with which may be associated distinctive organizing functions and motivations. (4) One could say with justification that Freud still intended to view the ego's functions in relation to reality mainly from the perspective of an environment that embodied actual and potential dangers and frustrations. Thus, the ego's functions in the

positive role of creator, rather than defender or adjuster, were only hinted at. (5) Finally, the issues of the nature and bases of the ego's autonomy were left hanging, as was the problem of how one might now conceptualize the ego's development independent of drive development. A reflection of this incompleteness was a curious ambiguity that remained in Freud's notion of the anxiety signal: the new conception of anxiety did not distinguish libidinal from nonlibidinal sources of anxiety, although it implied that there was such a distinction.

### Post-Freudian development

The concepts of the new ego psychology were amplified and systematized by Heinz Hartmann and supplemented by Ernst Kris, Rudolph Loewenstein, and David Rapaport, the torchbearers of the new ego psychology. Erik Erikson's efforts to delineate a biosocial conception of adaptational development are also pertinent, although he does not explicitly use an ego-psychological frame of reference. Erikson's work, in fact, highlights certain of the difficulties inherent in ego psychology's attempts to distinguish between drive "structure" and ego.

Hartmann perceived the central importance of Freud's assumption of an autonomous activity principle (he calls it the synthetic, or organizing, function of the ego) and made it the key feature of his conception of an ego system which undergoes a development that is coordinate and interactive with that of drives yet remains distinctive. Within Hartmann's scheme the ego emerges more clearly as a general adaptational "organ," reflecting in the range of its responsiveness responsibility to "realities" in the broad sense, as well as in the narrow sense of serving the pleasure principle. He views the ego in a problem-solving role, not simply as mediator of conflict; responsiveness to conflict becomes but one aspect of the ego's function in the adaptational task of "fitting in" with an "average expectable environment." Even the scope of conflict itself is greater than the clash of drive impulse and defense. As a self-regulative system, the ego responds to the requirements of different types of "disequilibria."

A central theme is Hartmann's conception of the autonomy of the ego's functions—that is, the degree of independence of ego activities from drives and from involvements in conflict. Phylogenetic development has ensured that the human child will come into the world with an array of functions potentially adaptive to an average expectable environment, provided adequate environmental opportunity is supplied. These functions include the basic



capacities and abilities of the organism: the abilities to perceive, to learn, to remember, to think, to move, and to act; talent and inborn gifts of all kinds, as well as dispositions to organize, to synthesize, to achieve balance from disequilibrium. The development of these functions is also relatively independent of drive, guided in large measure by the necessities of stimulation or contact with an environment that supports the actualizing of these potentials. (Relevant to this point is the great difficulty experimenters encounter in demonstrating consistent effects of needs on perception.) The assumption of conflict-free structures is extended even to functions that are responsive to parental and social realities. For instance, there is a conflict-free superego aspect of ego development. This framework makes it possible to accommodate such nonpsychoanalytic developmental theories as that of Piaget, for whom stages of the development of thinking are essentially the outcomes of interactions of perception and of action accompaniments of cognitive operations. [See DEVELOPMENTAL PSYCHOLOGY, *article on a THEORY OF DEVELOPMENT.*]

These assumptions of primary autonomy of functions signify recognition of a range of phenomena that do not owe their essential form to psychosexual influence. Indeed, the issue of the participation of drive and more generally of libido on ego development now becomes a more troublesome matter. Hartmann is careful to point out that "conflict-free" does not mean "drive-free" and that actually, many ego functions are molded on the pattern of such instinctual drive manifestations as giving and taking, introjection and projection, holding on and letting go; in these instances the ego functions have their analogue in the modes of motor functioning of the various erogenous zones.

For Erikson, this has been a departure point for developing a psychosocial conception of drives. For example, each of the zones that successfully attain a place of principal importance in the discharge of drive tension has one or more modes of operation that are basic for dealing with reality. The mouth lends itself to getting (i.e., receiving) and taking, the anus to holding on and letting go, and so forth. A critical point is that there takes place a process of estrangement, or autonomous separation, of the modes from their original locus in zones. They become, then, general ways of approaching the world. Erikson offers clinical examples from play configurations, dreams, neurotic symptoms, social behavior, etc. which show the utility of this way of viewing consistencies in behavior.

It should be pointed out, however, that Erikson's

mode-zone conception is actually more a reformulation of drive theory than an extension of the new ego conception of autonomy, for the core of his theory is that the kinds of functions associated with autonomous ego functions in the Hartmann scheme (e.g., perception, motor, etc.) are present as a part of drive structure from the very first. It is no solution—nor is it meant to be—to the dilemma created by Hartmann's notion of how much primary autonomy is to be allowed to ego. Erikson proposed that drives are always implicated in psychosocial adaptational modes, having a formative influence on the modes as well as being themselves shaped by the modes, but this seems to undermine the conception of a drive-autonomous development, in contradiction to Hartmann's view.

Conflict retains its importance in Hartmann's schema as an important condition for the emergence of new varieties of motivation and of action structures. There are two distinctly novel features, however, in his view of the relation of conflict to development. One is his proposal that structures of regulation can undergo a "change of function" from defense to adaptation. A regulative process that developed as a means of coping with conflict can be used for purposes other than defense; a defense such as intellectualization or denial may turn into a means of nondefensive adaptation. It becomes, as it were, a generalized mode of cognitive control suited to a range of encounters requiring resolution. These are examples of secondary autonomy in the sense that, although they originated from conflict in the service of drive control, they developed into effective instrumentalities of adaptation generally. Secondary autonomy is a reminder that the conflict-free realm of primary autonomous functioning can be enlarged by the development of new structures. The natural course of development is seen, then, to be enlargement of the conflict-free sphere, building on the base of the primarily autonomous structures additional structures that serve to contain the drives, slow them, and harness them to useful work.

A second novel feature of Hartmann's views has to do with alternative origins of defense itself. It is possible for the primary autonomous structures to play a role in predisposing the person toward certain modes of defense when he experiences conflict. In Freud's earlier views adaptive modes of control (usually termed character defenses) themselves originated in defensive reactions and/or in fixations on one or another mode of drive gratification. Now Hartmann is saying that characterological style and adaptive control may not only be fashioned from defenses but that the reverse is also possible

—defenses may be patterned on an autonomous behavioral tendency. Thus, a special talent may affect the distribution of libidinal and aggressive energies; it may itself be useful as a means of resolving conflict and, therefore, become a factor in determining habitual forms of defense. In this connection, the relationship between special sensitivities of infants and subsequent development of defenses in neurotic conflict is incorporated in this principle.

The developmental momentum of the conflict-free functions constitutes a source of independent motivation, of incentives for contacting the environment, and of adaptive control—a “third force,” on a par with drive and social reality. For instance, one source of motivation arises from the appearance, motivationally, of an anticipatory capacity which makes possible what Erikson calls time-locked motives. Hartmann attributes great importance to the anticipation function in ego development. Drives by themselves cannot guarantee adaptation to reality; neither is a heightened appreciation of reality enough to guarantee success for the ego's primary role nor renouncing an immediate pleasure enough to secure a greater and more certain one in the future. Neither are fear of pain nor the lure of pleasure, for that matter, in themselves precise indicators of direction and selection. It is necessary to assume a factor for anticipating the future, for orienting actions according to it, and for correctly relating means and ends to each other. This effort, if successful, yields pleasure secondarily; thus, integrative efforts become themselves a source of motives for contacting the environment and for developing interests and talents—none being derivable from drives directly.

Thus, Hartmann's formulation gives a place in psychoanalytic theory to function pleasure—that is, the intrinsically pleasurable aspects of the sheer exercise of functions—as a motivational source. Indeed, there is a hint that such motivations and the search for opportunities of exercising these functions without drive discharge as their basis are even necessary to normal development. Such intrinsic activity motivations guarantee, in Rapaport's view, the “stimulus nutriment” required for growth; Rapaport suggested that the seeking of stimulus nutriment is the basis of many activities described as “curiosity,” “search for novelty,” etc. Incidentally, another implication of these views is that growth itself and motives that stimulate activity to a great extent depend on the environment, not on the instinctual drives, in the sense that contact with the environment seems to be essential for the development and persistence of certain psy-

chological structures. The pleasures of such contact—and particularly of successful and effective contact—enlarge the organism's repertoire of motivations, beyond those originating in instinctual drives alone. [See STIMULATION DRIVES.]

It must be noted that in enlarging the picture of the sources of human motivation—motivations for dealing with reality that originate at least relatively independently of drives—Hartmann has also produced a more complex picture than heretofore of the potential conditions of, and participants in, conflict. For instance, motivations having to do with the function motives or intrinsic activity of the type mentioned above are not always consonant with those which derive from drives nor with internalized values. The disequilibria produced by conflict over drives, as well as those arising from within the ego system itself, become, then, a matter of psychoanalytic concern. Through his insistence on the importance of processes of adaptation to reality, Hartmann focuses attention on conflicts within, as well as between, systems: between drives and superego, between ego and superego, and between ego motives themselves.

The contemporary emphasis on the ego as a synthesizing system reflects a broader perspective of motivation than the earlier view, which saw it as a means of responding to drive tension. In the current view, the ego is a balance-inducing system, not simply a system for reducing tensions. Resolving the conflict of drive and superego demands is only one such task of the ego. The basic motivational conditions are still regarded as disequilibria, but these are not always assumed to implicate drive, nor do they always imply the reduction of tension; indeed, some conditions of balance and harmony may require an increase in the tension level of the organism.

Because the ego's synthesizing function is essential in modulating tension and achieving balance in a condition of disequilibrium, the ego is established as the prime—in fact, only—regulative system of the psychic apparatus to which the claims of drives and of superego are subordinated. The new ego psychology has irresistibly moved away from the tripartite model of control to a single control system in which the id and the superego are forces. The ego becomes the embodiment of all active purposive realignments of behavior. “Loss of control” in the new scheme connotes “ego deficiency.” While the dominance of the ego system is not an acknowledged precept of the Hartmann, Kris, and Rapaport conceptions, it is a logical implication of their views.

Ego functioning seen as a developmentally au-



onomous process extends psychoanalytic interest to the study of particular functions, such as perception and memory. A program for the analysis of thought functions encompasses the ways thought and behavior become alienated from reality, the conditions of conflict and drive investment that determine this alienation, and the behaviors that characterize it. For example, one might ask about the conditions under which, and the forms in which, perception develops adaptively in giving rise to effective psychical representations of reality; the conditions under which it may become sexualized or drawn into conflict; its role as helper or opponent of drives—that is, in its defensive, as well as drive-facilitative, forms; and its modification in relation to the equilibrating requirements of super-ego motivations.

Pathological phenomena can be conceptualized through consideration of conditions which produce a loss of “ego autonomy”—that is, a loss of the ego’s integrative capacity when drive and superego demands are beyond the organism’s tolerances. This derives from the view that ego autonomy is always relative; it is never absolute. Drives and their traces can invade thought organization and action.

In emphasizing the possibility that ego autonomy may be lost under conditions of increased drive pressure or weakened ego functions or, in Rapaport’s view, under conditions of an altered state of consciousness, provision is made for the fact that adaptations can take regressive, as well as progressive, forms. At the same time, the effort for adaptive synthesis and mastery (an ego motive) may come about in a temporary abeyance of ego control and may yield unique and novel forms of integration and insights into environmental realities—a conception which Kris brought into prominence while attempting to determine the conditions necessary for artistic creativity.

Two other contributions to the expanded significance given to the ego construct deserve mention. One is Rapaport’s attempt to reinstate consciousness as a decisive force in selection and action to a high place in the ego’s repertoire of regulative functions. While Freud emphasized the distinction between conscious and unconscious forces, the delineation of the functions of consciousness in adaptation and in the motivation of behavior was not one of his major interests. In the early period of the *Project*, consciousness was more or less identified with ego and was assigned adaptational importance, but later, when ego became virtually synonymous with defenses, consciousness receded in significance as a determining force of behavior.

However, the anxiety concept did reopen the general issue of the independently motivational importance of conscious states of feeling and affect and of modes of experience generally. The qualitative distinction between conscious and unconscious cognitive activity again assumed importance. For instance, Freud’s view that ego functions have different behavioral consequences in different states of consciousness exemplifies this concern.

Rapaport has described a variety of states and types of awareness and their behavioral effects. Perhaps the most important of his suggestions is the proposition that the less the capacities for vigilance, for reflective awareness, and for directing attention at will, the more do thought processes take on passive characteristics (i.e., show evidence of the primary process) and the less effective is secondary-process thinking.

### Ego psychology within the general theory

The radical reorientation within psychoanalysis brought about by the advent of ego psychology is perhaps most prominently exemplified in the development of psychoanalysis from a theory of pathology and illness to one that has ambitions of embracing all psychological disequilibria, whether specifically pathological or not. The change may be summarized as a shift from the conception of ego as an instrumentality of discharge of sexual and aggressive instinctual drives—and specifically as defense against those drives—to a concept of ego motivations derived from functions of adaptation and the resolution of crises. Several aspects of the changed emphases in the parent theory deserve comment.

**From conflict to crisis resolution.** Psychoanalytic theory resists single-trend interpretations of behavior; its propositions deal mainly with contrasting and incompatible forces, either actual or those of fantasy. It was the resolution of a particular form of opposition—the conflict between defense and drive—that occupied Freud in the main. This original insight into conflict as a source of motivation is now better regarded as a restricted expression of a more general view of life as a developmental process in which the resolution of incompatibility, crisis, and impasse is a critical condition of growth. The conflicts stressed by Freud are one aspect of a process of constant “adjudication” of incompatible polarities—for example, that we live in a phantom world of I–other relationships, whose code has to be understood by every person. Partly as a result of the impact of current ego psychology, the terms or polarities of potential impasse are more inclusive, encompassing a wide

variety of conflicts, including those forms not previously provided for in the theory. For example, Hartmann refers to conflicts within, as well as between, subsystems. Important emphasis is now given to developmental crises, in which modes of action and adaptation perfectly suited to the bio-social conditions of one stage are no longer appropriate to new maturational conditions and to social expectations tied in with these.

Crisis, impact, and dilemma produce a requirement for adjudication, and out of the variety of such felt or actual incompatibilities, including conflicts in the narrower sense of psychosexuality and defense, emerge motives of action and thought. The development of psychoanalysis has consisted in part of an even more discriminating appreciation of varieties of conflict and generally of the conditions of incompatibility that demand mastery. Psychoanalysis is still a psychology of motivation with emphasis on motives arising out of incompatibility, and this key principle is central to its clinical applications.

**From defense to adaptation.** It is also recognized that not all adaptational efforts are defense in the narrow sense. The emphasis on defense originated in a concern with illness and neurosis, the concept of conflict being the core of the theory. Now, as ego psychology becomes more explicitly a conception of adaptation and of developmental solutions, it is logical that it embody the assumption that adaptation is multiple faceted, with defense only one of its forms.

**From instinctual drives to motivational diversity.** The distinction between drive-originated and non-drive-originated motivations is increasingly a sensitive issue in contemporary psychoanalysis. The important clinical concern with motivational genesis—how the patient's goals came to be what they are—often turns out not to implicate psychosexuality. Clearly, not all motivations—in the sense of purposively instigated contact with the environment—can be referred to the main driving causes—sexuality and aggression. Confronting psychoanalytic theory is the problem of embracing the uneasy recognition of extrainstinctual motivations without repressing the hard-won realization that such drives as sexuality and aggression indeed have a plasticity not given to other sources of motivations—for example, displaceability of drive aims and objects, a not easily inhibited peremptoriness, and a higher potential for involvement in conflict. The strength and ubiquity of unconscious love and hate, including their self-directed aspects, must be incorporated into a comprehensive motivational theory in the future. Erikson's formulations are the

closest approximation we have at the present time. The capacity of traditional libido theory to effect this reconciliation is highly doubtful, and in fact, the uncritical acceptance of libido theory within the newer emphases of ego psychology sharpens the theoretical dilemma.

**From explanatory priority of unconscious wishes to explanatory parity of conscious purpose.** The polarity of consciousness-unconsciousness was always significant to Freud, but mainly in respect to consciousness as *reactive* to unconscious forces, particularly in his early formulation of the conscious-ego system. Mainly because of Rapaport's efforts, the causal importance of consciousness and of modes of experience is generally acquiring increasing respect in psychoanalytic theory. One aspect of this emphasis is a set of new possibilities opened up in respect to the importance of *affects*, previously considered as part of the anxiety-signal theory. One possibility is that an affective experience can be an independent source of motivation. The possibility has even been suggested that only via affect signals do drives themselves become motivations—that is, directional influences. At any rate, affect signals have been increasingly regarded as antecedents that determine realistic and unrealistic implementations of drive activation.

This change of emphasis in contemporary psychoanalysis is well illustrated by a reinterpretation of depression offered by Bibring to replace the traditional theory (1953). In the older theory, guilt over oral and aggressive strivings were assumed to be central in depression. In Bibring's theory, depression is considered to be primarily an ego phenomenon "representing an affective state of helplessness" which comes about because two incompatible sets of ego cognitions are simultaneously active: feelings of lowered self-esteem, on the one hand, and, at the same time, unrequited aspirations and a not yet relinquished sense of potential. Low self-esteem, inhibition of ego aspirations, and helplessness are the important focuses of the syndrome. Bibring actually assigns a peripheral role to the drive factors that are central in the accepted theory. Drives may precipitate or complicate depression or be exacerbated as a result of the *ego state*, but they are not primary determinants of depression, the crucial factor being the *ego state*. The theory is truly a radical reformulation in several respects: (1) It assigns priority, not to instinctual drives, but to an affective state of the ego. (2) It assigns importance to goals and motivations that are explicitly of the ego and not of drives. The incompatibility is "intrasystemic" and does not implicate the id. (3) It assumes con-



scious experience to be an active force in determining behavior, in asserting that the experience of helplessness is a key event in the depressive syndrome. (4) It draws attention to dilemma in the broad sense, rather than to conflict in the narrow sense, as a dynamic principle. (5) It accords a determining influence—even as a “fourth force”—to a grade of ego structure having to do with self-esteem. (6) Overlapping with the above, it regards instinctual sexual drives and aggression as complications, but not prior determinants, of the depressive state. [See DEPRESSIVE DISORDERS.]

The effects of the expanded conception of the psychoanalytic enterprise and of the newer emphasis on the ego have been both vitalizing and humbling. It has, for example, decisively checked the earlier pansexual reductionism with which psychoanalytic therapy and theory were burdened for years. Moreover, recent developments have brought a broadened appreciation of social and cultural influences into psychoanalytic theory. Psychoanalysis has come a long way from its beginnings, when the ego was the contravening agent against the id and against the culture, to appreciation of the fact that only a supporting society and a loving mother can make a functioning ego. Ego psychology attempts to correct the earlier de-emphasis, while avoiding a simple environmentalism.

At the same time ego psychology has been a vitalizing influence in making psychoanalysis aware of the processes by which reality is engaged, of the varying modes and styles of such adaptational approaches, and of the pathological consequences of a break in the equipment of effective engagement with the environment. This last aspect has been a particularly important spur to experimental work in nonclinical settings. Also prominent has been the burst of interest in conceptions of ego development, a field of study distinct from the study of cognitive development per se or of intelligence, and emotions.

Furthermore, psychoanalytic therapy no longer regards its main task to be the uncovering of unconscious memories. It is now more a matter of deciphering and revealing the pattern of object relationships and of specifying the scope of these patterns of relationship in behavior and experience. The goals of psychoanalysis have thus become more ambitious and more difficult, and the hopes of success, consequently, more modest.

Some psychoanalysts have cast a jaundiced eye on the current ego psychology, viewing it as a throwback to attempts to undo the importance of sexuality and aggression in behavior—to re-repress

the insights gained from Freud's long period of concern with instinctual drives. This is a misleading interpretation of the development. It seems more likely that it reflects an attempt to deal with motivations in behavior that cannot be easily explained in terms of libido theory. The explicit aim of ego psychology is not at all to pre-empt the importance of the hard-won clinical facts descriptive of conflict, defense, transference, regression, fixation, unconscious fantasy, and screen memories—all of which require proper understanding of the powerful forces of sexuality and hate—but to make them part of a larger framework which will encompass facts which these concepts are either not consistent with or insufficient to deal with.

### Evaluation of the concept

This seeming vitality and broadened perspective ought to be the happy note on which to end the present account. Ego psychology would appear to be a vigorous development on the psychoanalytic scene, one of increasing explanatory promise. Unfortunately, for all of the incentives to investigation promoted by ego psychology, there are grave doubts about its viability as theory in its present form.

A suspicious sign is the very inclusiveness of ego psychology. The ego has come to connote all those structures of man which, in the face of peremptory drive motivations, make him responsive to reality. Here is a summary of ego functions offered by two spokesmen for the new ego psychology: (1) consciousness; (2) sense perception; (3) the perception and expression of affect; (4) thought; (5) control of motor action; (6) memory; (7) language; (8) defense mechanisms and defensive activity in general; (9) control, regulation, and binding of instinctual energy; (10) the integrative and harmonizing function; (11) reality testing; and (12) the capacity to inhibit or suspend the operation of any of these functions and to regress to a primitive level of functioning (Arlow & Brenner 1964, p. 39). And this list does not even include the functions of the superego, which could quite properly be included as an ego “subsystem.” It is freely admitted, moreover, that ego theory offers no systematic principles by which to order its component functions hierarchically; which functions are most central and which most peripheral to ego as a structure are by no means clear. One result is vagueness in distinguishing ego development as a process from the development of intelligence, affect, and cognition generally and, indeed, from drive development.

Scientific theories have a way of hiding difficul-

ties by attempting to resolve impasses by complication. Certainly, current ego psychology complicates the picture of conflict, motivation, and determinism of behavior, compared to the earlier psychoanalytic conceptions. Conceptual complexity is no sure sign of better theory; indeed, it may well signify weakness and error. Some of the ambiguities and difficulties in current ego psychology therefore deserve scrutiny. Four unresolved issues in ego psychology cast doubt on the fundamental viability of the concept.

**Entity or taxonomy?** One difficulty in current formulations has to do with confusion over the appropriate data language to be used in designating ego phenomena. An aspect of this problem has to do with specifying the essential referent of the ego concept: whether the ego is a hypothetical unit or system of self-regulation with as yet unknown operational parameters, simply a taxonomical device for grouping functions according to their relation to id and superego, or a class of nonlibidinal motivations. Proponents of ego psychology are equivocal on this matter, and various theorists have even tried to dismiss the issue altogether. The first meaning—the ego as a regulative system—has definitely the ring of reification. In some accounts a definite structure, in some sense, of a hypothetical central mechanism is implied—structure in the “hard sense,” as in holistic references to the ego as “strong” or “weak” or to “autonomy of ego” from id.

On the other hand, reification is often and vigorously disclaimed; it is said that the ego’s “existence” is a matter of convenient metaphor (e.g., Waelder 1960). Something of the same type of usage, it is said, is seen in modern physics; an electron can be located only when it sends a message—that is, when it emits the ray in transmission from one energy level to another. And so it is with ego, superego, and id. The ego has its usefulness as an agent in sharpening the picture of the forces of conflict.

Despite such disclaimers, it is difficult to wave aside the possibility that the structural point of view carries inescapable assumptions of agency and process. In the early phases of his theorizing, Freud did not consider the ego simply as a grouping of functions; it had specifiable operational characteristics. Even if one takes the notion of the ego as a “group of functions”—that is, in its most popular current, taxonomical meaning—the very word “group” brings up questions of the processes of grouping, the nature of coherence, of why and how certain functions collaborate and others do not.

Moreover, avoiding the ego as a motivational organization—a meaning also encountered—carries with it the assumption of interlocking purposes and development, and again, the nature of the coherence becomes an issue. Loewald has made this point well. There is a distinction, he says, between processes and the structures between and within which subprocesses take place. He and Eissler raise the question of how the three structures of personality function. It is difficult to see how current ego psychology working from the “structural point of view” can avoid assuming some kind of operational model of a system working as a whole, especially in presenting itself as an explanatory theory of causation and in eschewing the more modest and achievable goal of being a descriptive psychology of meaning and purpose. But failing to confront its covert assumptions of a process model, the causal explanations offered by ego psychology are bound to appear to beg the question.

There is more than mere suspicion that the new ego psychology does indeed lean upon a model of hypothetical process without acknowledging it as such—namely, the tension-discharge-equilibrium model. Disclaiming reification, ego psychology, nevertheless, does implicitly hold to the pseudoneurological conception of psychical structure that is embodied in Freud’s energy conception of drives, and to the extent that it does, it inevitably has to reflect the limitations and inadequacies of this model of process and structure.

Some theorists (e.g., Rubenstein 1967) propose that ego theory ought to be directed toward the development of a model of self-regulation, one which would fill the gap in an uncertain metapsychology of mechanism, which the theory implies. Thus, there have been proposals to conceptualize autonomy in terms of some actual—perhaps physiological—model of information processing. This would put ego psychology squarely into the business of investigating central processes.

Most psychoanalytic theorists are understandably reluctant to move in these directions, for to do so would commit theorists to a data language not easily suited to the observables and aims of the clinical situation and would very likely even require data of a different order than are obtainable in the consulting room. More important are well-founded doubts as to whether such models could ever be productive to clinical work or could ever move beyond the stage of simply translating the purposivistic concepts of clinical work to a seemingly nonteleological cast. The challenge to those calling for psychoanalysis to move toward develop-



ing models of operational process is this: unless such models have built into them capacities for consciousness, modes of experience, and reflection on social responsibility, they are not likely to be of much use to the clinician.

One could try to wave aside the issue by democratic gesture: Let those who wish go the direction of mechanism; live and let live. This is well and good, but it requires the existence of an alternative course. It is doubtful, however, that the currently popular structural point of view provides such a clear alternative. Proponents of the structural point of view—certainly the prevalent one—shun the responsibility of developing the operations model they implicitly avow, in hopes that an explicit concern with the functions through which reality-adaptive efforts are carried out will suffice.

However, a serious contributor to theoretical ambiguity is the fact that several meanings of "function" are used interchangeably and confounded in contemporary ego psychology: "function" in the sense of "survival utility," "function" as "purpose" or "motivational aim," "function" with the non-motivational meaning of "capacity" (e.g., perception, memory).

Vagueness in the conception of functions affects the meaning and utility of the key concept of primary and secondary autonomy of functions. A common illustration of primary autonomy is the great extent to which perception, motility, and learning are relatively independent of conflict and drive. Here the word "function" clearly has the meaning of "capacity," "faculty," or "apparatus," without implication of "purpose." But, if "perception" and "motility" are not motivational terms, in what sense can they be autonomous, especially since perception can be an instrumentality both of ego and id motivations? Secondary autonomy almost always refers to a "change of function" in the purposive sense, rather than in the sense of "capacity." Defense, for instance, can become a generalized character trait. Thus, the seemingly unitary term "autonomy" lacks theoretical symmetry. The notions of primary and secondary autonomy are anchored to different meanings of "function," and in the theory what is presumed to be autonomous is ambiguous.

**Relationship to drive.** Ego psychology presents a paradox, rather than offering a solution, in assuming, on the one hand, that certain functions undergo a development outside of conflict, and, on the other, that "conflict-free" does not mean "drive-free." Autonomy means "autonomy from conflict," not "autonomy from drive."

Ego psychology seems not to have confronted the implication that if there is only one set of instrumentalities of perceiving, motility, etc., involved in the structuring of drives and, indeed, of all motivations, the necessity of assuming an agency distinct from id (and therefore two sets of instrumentalities, one for drives and one for ego motivations) becomes rather less urgent, even superfluous; if the distinct agency is retained, it becomes simply another and redundant way of referring to self-regulation in the organism as a whole, which includes drive as one aspect of its totality.

A key change in Freud's later conception of the ego was his recognition of an active motivational principle of growth that went beyond the pleasure principle of libidinal gratification—the turning of enforced responses into ego-initiated responses. The explication of this critical precept of ego psychology is still in its beginnings (Erikson's formulations are a most promising start). The unsettled question is whether there is a psychoanalytic version of ego development that is distinguishable from psychosexual development.

**Catch-all or distinctive theory?** In this sense, contemporary ego psychology is a reminder to psychoanalysis of theoretical tasks hardly begun. Merely to list the ego's functions is not to account for the logic of their grouping and the dynamic principles by which functions cohere and differentiate. What Hartmann and his collaborators have done in acknowledging conflict-free processes, as Loevinger points out, is to give ego functions "a local habitation and a name within systematic psychoanalysis" (1966, p. 433). This is an achievement, but it is not yet an integration of ego psychology into psychoanalysis. Acknowledging a "synthetic," or "integrative," function of the ego is not to state a dynamic principle by which to explain behavior in terms of synthesis activity. It is less important to assert that motivations can emerge from a process of "secondary autonomy" than to specify the dynamic principles that account for their origin and their subsequent predominance over developmentally earlier motivations. Lacking such dynamical principles of epigenesis, many of the concepts of ego psychology, such as primary or secondary autonomy, seem less like explanations than descriptive and classificatory formulations of phenomena that psychoanalytic theory had previously made it difficult to recognize.

The fundamental issue, then, confronting ego psychology is the question of whether the broad conception of the ego as an arbiter of reality is a parsimonious theory, which, under a few prin-

ciples, systematizes the seemingly diverse and unrelated purposes and motives observable in behavior, or is a means of developing *ad hoc* propositions to provide for observations that do not square with the theory of libidinal energy.

It should be pointed out that the discoveries of the pervasive influences of sexuality in pathology and the observations of developmental continuity in sexuality are the bases of, but are not identical with, the libido theory. The conception of a libido energy which is inherently directional and has an inherited developmental course is a special hypothesis designed to pull together diverse manifestations and developmental aspects of sensual pleasure. Thus, confirmation of the pervasiveness of sexuality is by no means confirmation of the libido theory. There is extraordinary resistance to making the separation, even though it has been many times demonstrated that the libido theory far from suffices to deal with many strands of evidence and observation of sexuality. Yet, these inadequacies forced Freud to revise his account of the nature of motivation, and these revisions are a prominent part of ego psychology.

To appreciate better the link between the ego conception and the libido theory, recall Freud's early model, which assumed a flow of energy into the mental apparatus from two peripheral sources, external and internal. Drive, the internal source, pervades and energizes the mental apparatus as an uncoded stimulus; "drive" thus had the dual connotation "stimulus and energizer." The ego concept was introduced to provide a responsible controlling agent, some means by which the uncoded stimulus is made cognitively sensible and given direction and which accomplishes and modulates the drive's "aim" of discharge. In libido theory, too, the notion of an energy which builds up tension, requires discharge, and is modulated to an equilibrium is fundamental. In the libido conception the notion of ego control came to place more emphasis on defense, variously described in terms of censor, repression, and even for a short time ego drive.

However, the challenge of how to conceive of the ego apparatus was a much more formidable one in the context of the libido theory than in the simplistic model of "endogenous" and "exogenous" energies of the *Project*. In the latter, drive was simply an internal energetic quantity. The ego concept could accommodate that. But, the libido theory proposes something more: it connotes an energy quantum having its own aims, range of objects of discharge, epigenesis, and even its own mecha-

nisms (e.g., displacement, consideration). In the later, tripartite model, id is truly an agency of impetus, requiring a very special agency of control. It gradually became evident that not all motivational phenomena could be encompassed by the libido conception. Freud faced a choice: Was the trouble to be diagnosed as a defect of the libido conception or as an inadequate picture of ego? Freud's decision was unequivocal. The conception of the ego as a controlling agent was changed, not the libido theory. Thus, a case can be made for the proposition that every change made thereafter in the ego conception was indirectly a supportive effort to retain the libido conception unaltered.

With expansion of the ego concept, however, the strains of coexistence have become even more evident. Ego theory in its late development has had to take account of the adaptive plasticity of behavior, cognitive autonomy from stimulation, capacities for choice and volition, and the independence of abilities and intelligence. But, the libido concept presents an impediment to such a picture of the ego; it assumes a distinctive quality of energy, having its own aims and structuring implementations independent of the ego and undergoing its own distinctive development, and requires an ego whose dominant function is to serve as a control over the claims for discharge of sexual and aggressive energies. This simply contradicts the trend toward a broadened conception of an ego as an autonomously motivating agent. The trouble began with accumulated evidence of reality-adaptive motivations that had no easily determined relationship to sexuality. It is to Hartmann's great credit that he confronted the motivational phenomena that forced extension of ego theory, but recourse to continuous enlargement of the ego concept is no strategy of parsimonious solutions.

The signs of strain are evident in the ingenious recourse to the concept of neutralized drive energy as a solution to this dilemma. Libido retains its status as an independent energetic force. However, this hardly clarifies how the allegedly independent drive energy can come under the sway of the structuring functions of the organism. There seems to be evidence that for sexuality to develop, the right psychosocial setting and adaptive instrumentalities must be present; in Rapaport's terms, drives also need their "appropriate supplies" of environmental stimulation. Solution of this dilemma still confronts contemporary psychoanalysis. [See *SEXUAL BEHAVIOR*.]

It seems likely that resolution of the difficulties in meshing libido theory and ego psychology will









- Freud. London: Hogarth; New York: Macmillan. → First published as *Zur Ätiologie der Hysterie*.
- FREUD, SIGMUND (1896b) 1962 Further Remarks on the Neuro-psychoses of Defence. Volume 3, pages 157–185 in *The Standard Edition of the Complete Psychological Works of Sigmund Freud*. London: Hogarth; New York: Macmillan. → First published as "Weitere Bemerkungen über die Abwehr-neuropsychosen."
- FREUD, SIGMUND (1900) 1953 *The Standard Edition of the Complete Psychological Works of Sigmund Freud*. Volumes 4–5: The Interpretation of Dreams. London: Hogarth; New York: Macmillan. → First published as *Die Traumdeutung*.
- FREUD, SIGMUND (1905) 1953 Three Essays on the Theory of Sexuality. Volume 7, pages 123–245, in *The Standard Edition of the Complete Psychological Works of Sigmund Freud*. London: Hogarth; New York: Macmillan. → First published as *Drei Abhandlungen zur Sexualtheorie*.
- FREUD, SIGMUND (1911) 1959 Formulations on the Two Principles of Mental Functioning. Volume 12, pages 215–226 in *The Standard Edition of the Complete Psychological Works of Sigmund Freud*. London: Hogarth; New York: Macmillan. → First published as *Formulierungen über die zwei Prinzipien des psychischen Geschehens*.
- FREUD, SIGMUND (1915) 1962 Instincts and Their Vicissitudes. Volume 14, pages 117–140 in *The Standard Edition of the Complete Psychological Works of Sigmund Freud*. London: Hogarth; New York: Macmillan. → First published as *Triebe und Triebschicksale*.
- FREUD, SIGMUND (1920) 1955 Beyond the Pleasure Principle. Volume 18, pages 7–66 in *The Standard Edition of the Complete Psychological Works of Sigmund Freud*. London: Hogarth; New York: Macmillan. → First published as *Jenseits des Lustprinzips*.
- FREUD, SIGMUND (1923) 1961 The Ego and the Id. Volume 19, pages 12–63 in *The Standard Edition of the Complete Psychological Works of Sigmund Freud*. London: Hogarth; New York: Macmillan. → First published as *Das Ich und das Es*.
- FREUD, SIGMUND (1926) 1959 Inhibitions, Symptoms and Anxiety. Volume 20, pages 77–178 in *The Standard Edition of the Complete Psychological Works of Sigmund Freud*. London: Hogarth; New York: Macmillan. → First published as *Hemmung, Symptom und Angst*.
- FREUD, SIGMUND (1932) 1962 New Introductory Lectures on Psycho-analysis. Volume 22, pages 3–158 in *The Standard Edition of the Complete Psychological Works of Sigmund Freud*. London: Hogarth; New York: Macmillan. → First published as *Neue Folge der Vorlesungen zur Einführung in die Psychoanalyse*. See especially pages 81–111, "Anxiety and Instinctual Life."
- FREUD, SIGMUND (1937) 1964 Analysis Terminable and Interminable. Volume 23, pages 216–253 in *The Standard Edition of the Complete Psychological Works of Sigmund Freud*. London: Hogarth; New York: Macmillan. → First published as *Die endliche und die unendliche Analyse*.
- FREUD, SIGMUND (1940) 1964 An Outline of Psycho-analysis. Volume 23, pages 141–208 in *The Standard Edition of the Complete Psychological Works of Sigmund Freud*. London: Hogarth; New York: Macmillan. → First published as "Abriss der Psychoanalyse."
- GARDNER, RILEY W. et al. 1959 *Cognitive Control: A Study of Individual Consistencies in Cognitive Behavior*. Psychological Issues, Vol. 1, No. 4. New York: International Universities Press.
- GILL, MERTON M. 1959 The Present State of Psychoanalytic Theory. *Journal of Abnormal and Social Psychology* 58:1–8.
- GILL, MERTON M. 1963 *Topography and Systems in Psychoanalytic Theory*. Psychological Issues 3, no. 2. New York: International Universities Press.
- GILL, MERTON M.; and KLEIN, GEORGE S. 1964 The Structuring of Drive and Reality: David Rapaport's Contributions to Psycho-analysis and Psychology. *International Journal of Psycho-analysis* 45:483–498.
- GLOVER, E. 1947 Basic Mental Concepts, Their Clinical and Theoretical Value. *Psychoanalytic Quarterly* 16: 482–506.
- GUNTHER, HARRY 1961 *Personality Structure and Human Interaction*. New York: International Universities Press; London: Hogarth.
- HARTMANN, HEINZ (1927–1959) 1964 *Essays on Ego Psychology: Selected Problems in Psychoanalytic Theory*. New York: International Universities Press.
- HARTMANN, HEINZ (1939) 1958 *Ego Psychology and the Problem of Adaptation*. Translated by David Rapaport. New York: International Universities Press. → First published as *Ich-Psychologie und Anpassungsproblem*.
- HARTMANN, HEINZ; KRIS, ERNST; and LOEWENSTEIN, RUDOLPH 1964 *Papers on Psychoanalytic Psychology*. Psychological Issues, Vol. 4, No. 2. New York: International Universities Press.
- HOLT, ROBERT R. 1960 Recent Developments in Psychoanalytic Ego Psychology and Their Implications for Diagnostic Testing. *Journal of Projective Techniques* 24:254–266.
- HOLT, ROBERT R. 1965 Ego Autonomy Re-evaluated. *International Journal of Psycho-analysis* 46:151–167.
- JACOBSON, E. (1953) 1960 The Affects and Their Pleasure-Unpleasure Qualities in Relations to Psychic Discharge Processes. Volume 1, pages 38–66 in Rudolph M. Loewenstein (editor), *Drives, Affects, Behavior: Contributions to the Theory and Practice of Psychoanalysis and Its Applications*. New York: International Universities Press.
- JONES, ERNEST 1957 *The Life and Work of Sigmund Freud*. Volume 3: The Last Phase 1919–1939. New York: Basic Books.
- KLEIN, GEORGE S. 1958 Cognitive Control and Motivation. Pages 87–118 in Gardner Lindzey (editor), *Assessment of Human Motives*. New York: Rinehart.
- KLEIN, GEORGE S. 1959 Consciousness in Psychoanalytic Theory: Some Implications for Current Research in Perception. *Journal of the American Psychoanalytic Association* 7:5–34.
- KRIS, ERNST (1939) 1952 On Inspiration. Pages 291–302 in Ernst Kris, *Psychoanalytic Explorations in Art*. New York: International Universities Press. → First published in Volume 20 in the *International Journal of Psycho-analysis*.
- KRIS, ERNST 1950a Notes on the Development and on Some Current Problems of Psychoanalytic Child Psychology. *Psychoanalytic Study of the Child* 5:24–46.
- KRIS, ERNST 1950b The Significance of Freud's Earlier Discussions. *International Journal of Psycho-analysis* 31:108–116.

- KRIS, ERNST 1951a Ego Psychology and Interpretation in Psychoanalytic Therapy. *Psychoanalytic Quarterly* 20:15-30.
- KRIS, ERNST 1951b The Development of Ego Psychology. *Samikšä* 5:153-168.
- KRIS, ERNST 1955 Neutralization and Sublimation: Observations on Young Children. *Psychoanalytic Study of the Child* 10:30-46.
- KRIS, ERNST 1956 The Recovery of Childhood Memories in Psychoanalysis. *Psychoanalytic Study of the Child* 11:54-88.
- LOEVINGER, JANE 1966 Three Principles for a Psychoanalytic Psychology. *Journal of Abnormal Psychology* 71:432-443.
- LOEWALD, HANS W. 1951 Ego and Reality. *International Journal of Psycho-analysis* 32:10-18.
- LOEWALD, HANS W. 1966 [A Book Review of] *Psychoanalytic Concepts and the Structural Theory*, by J. A. Arlow and C. Brenner. *Psychoanalytic Quarterly* 35:430-436.
- LOEWENSTEIN, RUDOLPH M. 1951 The Problem of Interpretation. *Psychoanalytic Quarterly* 20:1-30.
- RAPAPORT, DAVID 1951 Consciousness: A Psychopathological and Psychodynamic View. Pages 18-57 in *Conference on Problems of Consciousness*, Second, 1951. *Transactions: Problems of Consciousness*. New York: Josiah Macy Jr. Foundation.
- RAPAPORT, DAVID 1953 On the Psycho-analytic Theory of Affects. *International Journal of Psycho-analysis* 34:177-198.
- RAPAPORT, DAVID 1957 Cognitive Structures. Pages 157-200 in *Colorado, University of, Psychology Department, Contemporary Approaches to Cognition: A Symposium*. Cambridge, Mass.: Harvard Univ. Press.
- RAPAPORT, DAVID 1958 The Theory of Ego Autonomy: A Generalization. *Menninger Clinic, Bulletin* 22:13-35.
- RAPAPORT, DAVID 1959 A Historical Survey of Psychoanalytic Ego Psychology. Pages 5-17 in Erik H. Erikson, *Identity and the Life Cycle: Selected Papers*. Psychological Issues, Vol. 1, No. 1. New York: International Universities Press.
- RAPAPORT, DAVID 1960 On the Psychoanalytic Theory of Motivation. Volume 8, pages 173-247 in *Nebraska Symposium on Motivation*. Edited by Marshall R. Jones. Lincoln: Univ. of Nebraska Press.
- RAPAPORT, DAVID *The Collected Papers of David Rapaport*. Edited by Merton M. Gill. New York: Basic Books, 1967.
- RUBENSTEIN, B. R. 1967 Explanation and More Description: A Metascientific Examination of Certain Aspects of the Psychoanalytic Theory of Motivation. Pages 20-79 in Robert R. Holt (editor), *Motives and Thought: Psychoanalytic Essays in Memory of David Rapaport*. New York: International Universities Press.
- SHAKOW, DAVID; and RAPAPORT, DAVID 1964 *The Influence of Freud on American Psychology*. Psychological Issues, Vol. 4, No. 1. New York: International Universities Press.
- SPEERLING, SAMUEL J. 1958 On Denial and the Essential Nature of Defence. *International Journal of Psycho-analysis* 39:25-38.
- SPIZ, RENÉ A. 1960 *A Genetic Field Theory of Ego Formation: Its Implications for Pathology*. New York: International Universities Press.
- STEIN, M. 1965 States of Consciousness in the Analytic Situation, Including a Note on the Traumatic Dream. Volume 2, pages 60-86 in Max Schur (editor), *Drives, Affects, Behavior: Essays in Memory of Marie Bonaparte*. New York: International Universities Press.
- STONE, LEO 1962 *The Psychoanalytic Situation: An Examination of Its Development and Essential Nature*. New York: International Universities Press.
- WAEELDER, ROBERT 1936 The Principle of the Multiple Functions: Observations on Overdetermination. *Psychoanalytic Quarterly* 5:45-62.
- WAEELDER, ROBERT 1960 *Basic Theory of Psychoanalysis*. New York: International Universities Press.
- WHITE, ROBERT W. 1963 *Ego and Reality in Psychoanalytic Theory: A Proposal Regarding Independent Ego Energies*. Psychological Issues, Vol. 3, No. 3. New York: International Universities Press.

### III THERAPEUTIC METHODS

The psychoanalytic technique is a method of treating psychoneurosis which is based on uncovering the forces and motives that underlie the patient's symptoms. It originated in a chance discovery made in 1882 by Josef Breuer while treating a severely disturbed hysteric (the famous case of "Anna O."). When she was in a hypnotic state, this patient would talk to Breuer either spontaneously or in answer to his questions and would recount painful, upsetting events of her life that appeared to relate to elements of her symptoms or to words she would utter while in twilight states. The patient herself called this treatment "chimney-sweeping," because each recall of a traumatic event was followed by the disappearance of symptoms related to that event. For instance, her inexplicable terror of drinking water from a glass vanished after she remembered that she had once seen her governess let a lap dog drink out of a glass and that the sight had aroused considerable disgust in her.

#### Freud's development of psychoanalytic therapy

In 1889 Freud began to apply this *cathartic method*, as it was called, to other hysterical patients and was able to confirm Breuer's discovery. Freud and Breuer described their experiences in their *Studies on Hysteria* (1893-1895). In a later work (1914a), Freud made a record of some additional observations concerning the technique evolved by him. He stressed that for a satisfactory therapeutic result it was essential that the patient remember all traumatic events: the earliest and the most recent, as well as all intermediate ones; moreover, the remembering of traumatic situations had to be accompanied by the re-experiencing and verbal discharge of the affects connected with the



traumata, affects that had been "strangulated." [See HYPNOSIS; HYSTERIA.]

As he continued to treat patients in this way, Freud met with two shortcomings of the cathartic method. Some patients could not be hypnotized; others, although susceptible to hypnosis, experienced only temporary relief of their symptoms. In brief, he found hypnosis an unreliable therapeutic tool. For a short time he experimented with what he called the *method of concentration*, which consisted in prompting the patient to remember the forgotten events while pressing his hand on the patient's forehead. He described how the patient would respond to this prompting with ideas that occurred to him (free associations), which were either the looked-for traumatic events or some other experiences or thoughts or fantasies or dreams. In these seemingly unrelated associations, Freud soon learned to discern connections, allusions, and hints referring to symptoms, on the one hand, and to some as yet unremembered traumatic situations, on the other. Gradually he became able to guess or to infer, from all this material, the traumata on which the neurosis was based. At this point he developed the framework of what later was to be called the *method of interpretation and reconstruction* (1900; 1916–1917; 1937a; 1937b). At the time he developed this method, Freud made another discovery which was to become fundamental for psychoanalytic technique. His patients had great difficulties in remembering and were reluctant to communicate even seemingly innocuous ideas. Freud concluded that these inner obstacles—*resistances*, as he called them, borrowing a term from physics—were in fact the counterpart of the forces that kept the relevant memories and related affects shut off from consciousness.

The actual psychoanalytic technique, as we know it today, started when Freud replaced the method of concentration with that of *free association*. This technique has progressed throughout the past decades, but its essential features were laid down by Freud between 1900 and 1915 (1905a; 1908; 1909a; 1909b; 1911–1915; 1941a). Freud's gradual development of the method over those years resulted in successive changes in his theoretical views on the etiology and pathogenesis of neuroses.

Freud's early ideas on the nature of pathogenic traumatic events were epitomized in his theory of sexual seduction during childhood. The realization that many such sexual traumata had not really occurred, but represented the patient's unconscious fantasies, opened the way to his first discovery of *childhood sexuality* (1905b). Neu-

rotic symptoms were recognized as determined not merely by some fortuitous traumatic event but by an interplay of forces: by pathogenic conflicts between warded-off (repressed) instinctual demands and warding-off (defensive) forces of the ego. The psychoanalytic method strives to *make conscious* both the warded-off and the warding-off *unconscious* elements involved directly or indirectly in the patient's symptoms. It leads to therapeutic results on the basis of this transformation. The investigative and therapeutic aspects of the psychoanalytic method overlap partly, but not to the same extent in each case.

**Free association.** The *method of free association* expects the patient in analysis to say whatever occurs to him, without "censoring." This is called the *basic rule* of psychoanalysis. The method leads to a gradual unfolding of *preconscious* derivatives of the conflicting forces. In practice, this process is initiated by a compact made between the psychoanalyst and his patient: the patient promises complete candor; the analyst, total discretion. That is how Freud (1940) defined the *psychoanalytic situation*. The practical arrangements, consisting of regular and frequent treatment sessions, commonly of fifty minutes' duration and held four to six times per week, with the patient lying on a couch and the analyst seated in a chair behind him, have valid technical reasons. The psychoanalytic process unfolds in the most favorable way under these conditions.

The rationale for the method of free association is based on the observation that when conscious, goal-directed thinking is partly relinquished in favor of spontaneously emerging thought contents, part of the habitual defense against preconscious and unconscious material can be bypassed. The analysand's supine position tends to eliminate or weaken a further source of automatic censorship; it facilitates the verbalization of thoughts which would automatically be left unspoken in the usual face-to-face conversation. The result may be that the patient remembers situations he had forgotten, or expresses thoughts that at first seem foreign to him, or that familiar memories appear unexpectedly in a new psychological context. Or again, during some session a number of thoughts and memories may emerge in which the analyst will discern a common theme or denominator. The frequency and relative length of psychoanalytic sessions promote a gradual reactivation of past emotions, fantasies, memories, and thoughts, which are thus made available to the conscious ego of the patient.

The initial psychoanalytic situation develops into a process of interaction between analyst and patient in which each has a task to perform. The patient gradually learns to comply with the basic rule. The analyst assists him in doing so, learns to understand him, and communicates this understanding to him. It is a lengthy and highly complex procedure, mainly because some derivatives of pathogenic conflicts are, or become, involved in the process of free association itself and interfere with the patient's ability to follow the basic rule.

*Resistance and transference.* The way in which a given analysand observes the basic rule results in a variety of individual modes of free associating. These diverse modes are influenced, above all, by the particular resistances against the basic rule which inevitably emerge in every psychoanalytic treatment.

Despite sincere efforts to be candid, patients cannot avoid censoring some of their thoughts. Either they are ashamed or feel guilty about certain actions or wishes, or what occurs to them seems too irrelevant or stupid, too indiscreet or too offensive, to be uttered. At times, they go to any length to avoid expressing such a thought; or they recount it but omit some detail that is essential for its understanding. Or again, they communicate important thoughts or memories in a clinical and impersonal fashion, without being aware of, or able to express, the corresponding affect. And sometimes, although they feel they have very much to tell, they find it impossible even to utter a single word. These inner obstacles (resistances) to free association represent effects of the defensive forces, transferred from the pathogenic conflicts as such onto the psychic material which might reveal them during treatment. The unconscious struggle against bringing the warded-off pathogenic conflicts to consciousness reveals the underlying, unconscious, individually variable trend to preserve the neurosis.

Often the thoughts which patients are most reluctant to express have to do with the person of the analyst. The material of the free associations is connected either with the neurosis or with reactions to the analyst, and these reactions, feelings, and general attitude of the patient toward the analyst are so important that the outcome of the treatment may depend upon them. Freud took an essential step in elaborating the psychoanalytic technique when he recognized them as *transferences*, i.e., as transpositions from an infantile object onto the analyst or as modified unconscious repetitions, in the treatment situation, of early

warded-off tendencies and experiences, but now with the analyst as object.

This objective approach in dealing with the patient's reactions to the analyst has become an essential tool in analyzing those resistances which are of particularly far-reaching influence on the treatment: the so-called *transference resistances*. These may manifest themselves as direct obstacles to free association or may influence the associations in more subtle ways. For instance, the basic rule may be made inoperative by being taken too literally; it may be used for concealed parody of the analytic treatment or used in various indirect ways to gratify complex, unconscious transference reactions. The diverse transference resistances comprise old defenses and the instinctual demands themselves. Thus they represent a new version of infantile pathogenic conflicts unconsciously relived in the patient's efforts to associate.

*The genetic approach.* The fact that the psychoanalyst understands the patient's reactions toward him as manifestations of transference, and analyzes them from a dynamic as well as an ontogenetic point of view, is referred to as the *genetic approach* (Hartmann & Kris 1945). It calls for the use of genetic interpretations and of reconstructions of the patient's warded-off past experiences. This genetic approach is particularly prominent in the analyst's work when dealing with those phenomena which constitute a *transference neurosis* or, better, *transferred neurosis*; that is, when transferences are no longer isolated manifestations, but a considerable part of the patient's symptoms becomes involved with the person of the analyst or the process of analysis itself.

The work of interpreting becomes effective mainly when applied under favorable conditions and in particular ways. An important aspect of the analyst's work consists in working through the resistances which oppose progress and success of the treatment (Freud 1914a; 1914b).

When the patient has difficulties in following the basic rule of free association, the task of the analyst is twofold. He must assist his patient in overcoming these obstacles. On the other hand, he must understand, and help the patient to understand, what they reveal with regard to the neurosis. For example, if he becomes aware that the patient's free associating is hampered by an effort to avoid the expression of hostile thoughts against him, he will point out this conflict and will strive to understand and to interpret it as a modified repetition of similar situations that the patient has encountered with other persons. These actions of



the analyst further the process of free association, and at the same time affect the patient in a therapeutic sense.

**Interpretation.** The analyst has at his disposal a number of possible interpretations that are characteristic and specific for the psychoanalytic method. In psychoanalysis the term "interpretation" is applied to those explanations which add to the patient's knowledge of what has been warded off in himself. Interpretations aim at both the defensive and the instinctual motivational aspects of the conflict. In the example mentioned above, the interpretation can stress the struggle against aggression (*resistance analysis*); it can stress the aggressive (instinctual) character of the warded-off thought; ultimately, it should encompass both in their genetic perspective.

Although Freud had early emphasized the crucial importance of the analysis of resistances (1912a; 1912b), his major scientific interest for a long time focused on the study of instinctual drives. The psychoanalytic investigation of his patients' pathogenic conflicts led him to important discoveries concerning the repressed (or otherwise warded-off) instinctual impulses. These could be understood as derivatives of the libidinal and aggressive drives. He thus was able to reconstruct the maturational and developmental sequences of those drives in early childhood, as well as the specific types of object relations that correspond to their various developmental stages. The Oedipal phase of development is the best-known example. [See PERSONALITY, article on PERSONALITY DEVELOPMENT.]

In the course of the treatment, the existence of these drives, and their impact upon the patient's behavior in the past and in the present, are pointed out to him by interpreting his dreams, associations, symptoms. This is often referred to as *id interpretation*. However, these interpretations alone do not generally enable the patient to become conscious of the drive derivatives in a genuine, therapeutically effective way. They must therefore be preceded or supplemented by an analysis of his resistances, which makes the patient's adult ego better capable of tolerating his hitherto warded-off drives. The therapeutic objective of psychoanalysis is to change the balance of psychic forces that prevailed when the pathogenic conflicts were formed to a new balance better adapted to the patient's adult condition. Both the warded-off and the defensive forces are inaccessible to change so long as they remain unconscious. Hence the importance of analyzing the latter as well and making them conscious.

## Elaborations on Freud's initial contribution

Since Freud laid down the principles of psychoanalytic technique, our knowledge both of normal psychic functioning and of mental illness has grown considerably, and progress in psychoanalytic technique proper has been made. Many analysts have contributed to this technical progress and refinement. Of the many, I should like to mention only a few: Freud, in his papers "Analysis Terminable and Interminable" (1937a) and "Constructions in Analysis" (1937b); Ferenczi (1908–1933); Ferenczi and Rank (1924); Abraham (1907–1925); Nunberg (1920–1942). Other books are devoted entirely to psychoanalytic technique: Glover (1928); Lorand (1946); Menninger (1958); Nacht (1950); Alexander and French (1946).

**Ego psychology.** A particularly important development in psychoanalysis was the introduction of ego psychology. The scientific basis for a systematic, effective analysis of resistances (defenses) had been lacking before Freud's reformulation, in the 1920s, of certain fundamental propositions. The new theoretical concepts he advanced at that period laid the ground for a psychoanalytic *ego psychology*. Most important for the progress of analytic technique were his new formulations referring to *unconscious defenses* of the ego, to unconscious self-critical or self-punitive forces (the *superego*), and to anxiety as a signal operating as a trigger of defenses (1923; 1926).

Technical innovations based on Freud's theoretical reformulations were introduced by Wilhelm Reich (1933) and in a particularly fruitful way by Anna Freud. The application of ego psychology to psychoanalytic technique was further elaborated by such authors as Fenichel (1941); Hartmann (1951); Hoffer (1954); Kris (1951; 1956a; 1956b; 1956c); Loewenstein (1951a; 1951b; 1954; 1956; 1957; 1958; 1963); Nacht (1951); Stewart (1961).

The importance of the defenses of the ego has influenced the analyst's approach to the patient's free association. In the treatment situation the analyst thus pays attention not only to the material emerging when the patient follows the basic rule but equally to his struggles against it (Anna Freud 1936). The latter tend to reveal the characteristics of particular mechanisms of defense. For instance, whether a patient reacts to a hidden aggressive impulse against the analyst with fear of punishment, with exaggerated solicitude, with sudden silence, or in some other way may disclose a great deal about the character of his relevant defenses against aggression, not only in the present but also

in the past, and hence shed light on their role in his pathogenic conflicts. Thorough understanding of an individual patient's defensive mechanisms, and of their impact upon his neurosis, his general behavior, and the course of his analysis, is as important as the understanding of his instinctual drives. [See DEFENSE MECHANISMS.]

The functions of the ego are not limited to defensive mechanisms; some are relatively independent of the disturbing influence of drive derivatives. This has an important bearing on psychoanalytic technique, for the patient's ability to associate, to recover childhood memories, to gain insight, rests upon the *autonomy* of the relevant ego functions (Hartmann 1939). Their roles in the psychoanalytic process are of paramount significance (Hartmann 1951; Kris 1951; 1956a; 1956b; Loewenstein 1951a; 1951b; 1957; 1963).

**Sources of resistance.** In Freud's earliest formulation, the therapeutic effect of psychoanalysis had been attributed to the recall of forgotten traumatic events and to the abreaction of strangled affects. That assumption has long since been discarded. Current psychoanalytic theory sees the major therapeutic agent in the overcoming of resistances and in the bringing to consciousness of warded-off material: recovery of memories, reestablishment of disrupted connections, gaining of insight.

Psychoanalytic treatment must overcome many obstacles in order to achieve satisfactory results (Bibring 1937). Resistances to cure originate in the three "systems" of the psychic apparatus—the ego, the id, and the superego—and in external reality as well.

Rigidity, primitiveness, and the relative strength of defensive mechanisms represent the major part of observable resistances. Also, in most instances, the patient's ego derives advantage from his neurosis (*epinosis*, or "secondary" gain), causing him to cling to it, which can be a serious hindrance (Freud 1926). Such disturbances of specific autonomous ego functions impair analyzability (Hartmann 1951).

The strength and tenacity of pregenital components of the instinctual drives and, conversely, the weakness of their genital components constitute barriers to cure. These quantitative factors in the id must be considered in relation to the strength of the ego and not in terms of absolute values (Freud 1937a; 1937b).

Unconscious self-punitive demands of the superego often play a decisive role among the resistances to cure. While somatic illness usually

activates or complicates neurosis, it produces the opposite effect in a category of patients known as moral masochists. Their neurotic symptoms disappear when they are affected by a somatic disease. Once well again physically, they relapse into their neurosis. In these cases one can best observe the so-called *negative therapeutic reaction* (Freud 1923).

Social, cultural, and economic conditions can contribute to neurotic illness in some cases, to favorable conflict resolution in others. Familial circumstances especially may hold the balance in this respect. The momentous impact of parental behavior upon psychic development in childhood is well known. But also the presence or absence of neurotic disorder in other members of the family, or in the patient's husband or wife, may prove decisive for the outcome of an analytic treatment.

The recent advances in our knowledge, briefly outlined above, have not changed the technique of psychoanalysis itself. Rather, they have led to some shifts of emphasis and considerable refinements in its use. These modifications have broadened the applicability of psychoanalytic treatment and have improved its efficacy. In many instances, they can make the difference between failure and success.

Originally, psychoanalysis was found effective exclusively in the treatment of symptom neuroses in adults. Its successful use in treating the neuroses of children and adolescents opened up an important area for psychoanalytic therapy. Except for some changes in the procedure, e.g., substitution of the play technique for free association, the method in child analysis is basically the same as in the treatment of adults (see Anna Freud 1936).

The application of psychoanalysis to an ever greater variety of behavioral disorders, even to psychoses, produce striking results but also a number of disappointing failures. Many patients, and some analysts as well, set their aims too high. They expect psychoanalysis to achieve not merely a cure of neurotic illness, or even the emergence of a new and better personality, but—above all—lasting happiness for the patient. These hopes have obviously not been realized.

The widening scope of analytic treatment led, in its turn, to a greatly increased use of psychoanalytically oriented psychotherapy. The dividing line may be tenuous at times, but most forms of psychotherapy differ technically from psychoanalysis in rationale as well as in practical use. However, virtually all contemporary forms of psychotherapy draw to some degree upon psychoanalysis,



either as a method or as a body of knowledge. Psychoanalysis has enriched and profoundly influenced the entire fields of psychiatry and psychology.

RUDOLPH M. LOEWENSTEIN

[Directly related is the entry INTERVIEWING, article on THERAPEUTIC INTERVIEWING. Other relevant material may be found in CLINICAL PSYCHOLOGY; MENTAL DISORDERS, TREATMENT OF, especially the article on PSYCHOLOGICAL TREATMENT; NEUROSIS; PSYCHIATRY; and in the biographies of ABRAHAM; ALEXANDER; FERENCZI; FREUD; RANK; REICH.]

#### BIBLIOGRAPHY

- ABRAHAM, KARL (1907-1925) 1953-1955 *Selected Papers*. 2 vols. New York: Basic Books. → Papers and essays translated from the German, including Abraham's more important psychoanalytic works.
- ALEXANDER, FRANZ; and FRENCH, THOMAS M. 1946 *Psychoanalytic Therapy: Principles and Application*. New York: Ronald.
- BIBRING, EDWARD 1937 Versuch einer allgemeinen Theorie der Heilung. *Internationale Zeitschrift für Psychoanalyse* 23:18-37.
- FENICHEL, OTTO 1941 *Problems of Psychoanalytic Technique*. New York: Psychoanalytic Quarterly.
- FERENCZI, SÁNDOR (1908-1914) 1952 *First Contributions to Psycho-analysis*. London: Hogarth. → A selection of Ferenczi's essays originally published in German. First published in English as *Contributions to Psycho-analysis*.
- FERENCZI, SÁNDOR (1908-1925) 1927 *Further Contributions to the Theory and Technique of Psycho-analysis*. New York: Liveright. → A collection of some of Ferenczi's most important contributions to psychoanalysis. Translated from the German.
- FERENCZI, SÁNDOR (1908-1933) 1955 *Final Contributions to the Problems and Methods of Psycho-analysis*. London: Hogarth. → Contains Ferenczi's papers, translated from the German, that were written after 1926, as well as papers, fragments, and notes published posthumously.
- FERENCZI, SÁNDOR; and RANK, OTTO (1924) 1925 *The Development of Psychoanalysis*. New York: Nervous and Mental Disease Monographs. → First published as *Entwicklungsziele der Psychoanalyse*.
- FREUD, ANNA (1936) 1957 *The Ego and the Mechanisms of Defense*. New York: International Universities Press. → First published as *Das Ich und die Abwehrmechanismen*.
- FREUD, SIGMUND (1895-1926) 1953- The *Standard Edition of the Complete Psychological Works of Sigmund Freud*. 24 vols. London: Hogarth; New York: Macmillan. → First published in German. Twenty-three volumes have been published as of 1967.
- FREUD, SIGMUND (1897-1902) 1954 *The Origins of Psychoanalysis: Letters to Wilhelm Fliess, Drafts and Notes, 1887-1902*. New York: Basic Books. → Translated from the German.
- FREUD, SIGMUND (1900) 1953 *The Interpretation of Dreams*. Volumes 4-5 in *The Standard Edition of the Complete Psychological Works of Sigmund Freud*. London: Hogarth; New York: Macmillan.
- FREUD, SIGMUND (1905a) 1953 Fragment of an Analysis of a Case of Hysteria. Volume 7, pages 3-122 in

*The Standard Edition of the Complete Psychological Works of Sigmund Freud*. London: Hogarth; New York: Macmillan.

- FREUD, SIGMUND (1905b) 1953 Three Essays on the Theory of Sexuality. Volume 7, pages 123-245 in *The Standard Edition of the Complete Psychological Works of Sigmund Freud*. London: Hogarth; New York: Macmillan.
- FREUD, SIGMUND (1908) 1955 From the History of an Infantile Neurosis. Volume 17, pages 3-122 in *The Standard Edition of the Complete Psychological Works of Sigmund Freud*. London: Hogarth; New York: Macmillan.
- FREUD, SIGMUND (1909a) 1955 Analysis of a Phobia in a Five-year-old Boy. Volume 10, pages 3-149 in *The Standard Edition of the Complete Psychological Works of Sigmund Freud*. London: Hogarth; New York: Macmillan.
- FREUD, SIGMUND (1909b) 1953 Notes Upon a Case of Obsessional Neurosis. Volume 10, pages 151-249 in *The Standard Edition of the Complete Psychological Works of Sigmund Freud*. London: Hogarth; New York: Macmillan.
- FREUD, SIGMUND (1911-1915) 1953 Papers on Technique. Volume 12, pages 85-173 in *The Standard Edition of the Complete Psychological Works of Sigmund Freud*. London: Hogarth; New York: Macmillan.
- FREUD, SIGMUND (1912a) 1953 The Dynamics of Transference. Volume 12, pages 97-108 in *The Standard Edition of the Complete Psychological Works of Sigmund Freud*. London: Hogarth; New York: Macmillan.
- FREUD, SIGMUND (1912b) 1953 Recommendations to Physicians Practising Psycho-analysis. Volume 12, pages 109-120 in *The Standard Edition of the Complete Psychological Works of Sigmund Freud*. London: Hogarth; New York: Macmillan.
- FREUD, SIGMUND (1914a) 1957 On the History of the Psycho-analytic Movement. Volume 14, pages 7-66 in *The Standard Edition of the Complete Psychological Works of Sigmund Freud*. London: Hogarth; New York: Macmillan.
- FREUD, SIGMUND (1914b) 1953 Remembering, Repeating and Working-through (Further Recommendations on the Technique of Psycho-analysis II). Volume 12, pages 145-156 in *The Standard Edition of the Complete Psychological Works of Sigmund Freud*. London: Hogarth; New York: Macmillan.
- FREUD, SIGMUND (1916-1917) 1953 Introductory Lectures on Psycho-analysis. Part III: General Theory of the Neuroses. Volume 16 in *The Standard Edition of the Complete Psychological Works of Sigmund Freud*. London: Hogarth; New York: Macmillan.
- FREUD, SIGMUND (1923) 1961 The Ego and the Id. Volume 19, pages 12-63 in *The Standard Edition of the Complete Psychological Works of Sigmund Freud*. London: Hogarth; New York: Macmillan.
- FREUD, SIGMUND (1926) 1959 Inhibitions, Symptoms and Anxiety. Volume 20, pages 77-178 in *The Standard Edition of the Complete Psychological Works of Sigmund Freud*. London: Hogarth; New York: Macmillan.
- FREUD, SIGMUND (1937a) 1959 Analysis Terminable and Intermittent. Volume 5, pages 316-357 in *Sigmund Freud, Collected Papers*. International Psychoanalytic Library, No. 10. New York: Basic Books; London: Hogarth. → First published as *Die endliche und die unendliche Analyse*.

- FREUD, SIGMUND (1937b) 1959 *Constructions in Analysis*. Volume 5, pages 358-371 in Sigmund Freud, *Collected Papers*. International Psycho-analytic Library, No. 10. New York: Basic Books; London: Hogarth. → First published in German.
- FREUD, SIGMUND (1940) 1949 *An Outline of Psychoanalysis*. New York: Norton.
- FREUD, SIGMUND (1893-1895) 1955 *The Standard Edition of the Complete Psychological Works of Sigmund Freud*. Volume 2: *Studies on Hysteria*, by Josef Breuer and Sigmund Freud. London: Hogarth; New York: Macmillan.
- GLOVER, EDWARD (1928) 1955 *The Technique of Psychoanalysis*. New York: International Universities Press.
- GREENACRE, PHYLLIS 1956 Re-evaluation of the Process of Working Through. *International Journal of Psychoanalysis* 37:439-444.
- HARTMANN, HEINZ (1939) 1958 *Ego Psychology and the Problem of Adaptation*. New York: International Universities Press. → First published in German in the *Internationale Zeitschrift für Psychoanalyse und Imago*.
- HARTMANN, HEINZ 1951 Technical Implications of Ego Psychology. *Psychoanalytic Quarterly* 20:31-43.
- HARTMANN, HEINZ; and KRIS, ERNST 1945 The Genetic Approach in Psychoanalysis. *Psychoanalytic Study of the Child* 1:11-30.
- HOFFER, W. 1954 Defensive Process and Defensive Organization: Their Place in Psycho-analytic Technique. *International Journal of Psychoanalysis* 35:194-198.
- KRIS, ERNST 1951 Ego Psychology and Interpretation in Psychoanalytic Therapy. *Psychoanalytic Quarterly* 20:15-30.
- KRIS, ERNST 1956a On Some Vicissitudes of Insight in Psycho-analysis. *International Journal of Psychoanalysis* 37:445-455.
- KRIS, ERNST 1956b The Recovery of Childhood Memories in Psychoanalysis. *Psychoanalytic Study of the Child* 11:54-88.
- KRIS, ERNST 1956c The Personal Myth: A Problem in Psychoanalytic Technique. *Journal of the American Psychoanalytic Association* 4:653-681.
- LOEWENSTEIN, RUDOLPH M. 1951a The Problem of Interpretation. *Psychoanalytic Quarterly* 20:1-14.
- LOEWENSTEIN, RUDOLPH M. 1951b Ego Development and Psychoanalytic Technique. *American Journal of Psychiatry* 107:617-622.
- LOEWENSTEIN, RUDOLPH M. 1954 Some Remarks on Defences, Autonomous Ego and Psycho-analytic Technique. *International Journal of Psychoanalysis* 35:188-193.
- LOEWENSTEIN, RUDOLPH M. 1956 Some Remarks on the Role of Speech in Psycho-analytic Technique. *International Journal of Psychoanalysis* 37:460-468.
- LOEWENSTEIN, RUDOLPH M. 1957 Some Thoughts on Interpretation in the Theory and Practice of Psychoanalysis. *Psychoanalytic Study of the Child* 12:127-150.
- LOEWENSTEIN, RUDOLPH M. 1958 Remarks on Some Variations in Psycho-analytic Technique. *International Journal of Psychoanalysis* 39:202-210.
- LOEWENSTEIN, RUDOLPH M. 1963 Some Considerations on Free Association. *Journal of the American Psychoanalytic Association* 11:451-473.
- LORAND, SÁNDOR 1946 *Technique of Psychoanalytic Therapy*. New York: International Universities Press.
- MENNINGER, KARL A. 1958 *Theory of Psychoanalytic Technique*. New York: Basic Books.
- NACHT, SACHA 1950 *De la pratique à la théorie psychanalytique*. Paris: Presses Universitaires de France.
- NACHT, SACHA 1951 Les nouvelles théories psychanalytiques sur le moi et leurs répercussions sur l'orientation méthodologique. *Revue française de psychanalyse* 15:569-576.
- NUNBERG, HERMANN (1920-1942) 1948 *Practice and Theory of Psychoanalysis: A Collection of Essays*. New York: Nervous and Mental Disease Monographs.
- REICH, WILHELM (1933) 1961 *Character-analysis*. 3d ed., enl. New York: Noonday Press. → First published in German.
- STEWART, WALTER 1961 The Development of the Therapeutic Alliance in Borderline Patients. *Psychoanalytic Quarterly* 30:165-166.

## IV

## EXPERIMENTAL STUDIES

The impact of Sigmund Freud upon psychology and the social sciences generally has been very great. Freud is to be compared with other great figures in the history of science over the last century such as Darwin, Pavlov, and Einstein. His observations and discoveries about people, brilliantly set forth, assure him a place in history regardless of the changes that may be wrought in the theories that he proposed. The clarification of the place of psychoanalytic facts and theories within the social sciences must rest eventually on their absorption into the social sciences after they have been reconstructed and validated in their new form. Such reconstruction and validation can take place by applying the methods of empirical study, model building, and hypothesis testing already familiar within these fields, or perhaps by new methods invented to deal with the subtleties of the clinical interaction between analyst and patient.

## Historical background

Psychoanalysis has a special history because it was dominated for such a long time by Sigmund Freud, who began theory construction at the very start of his career and fought the social battles for recognition. There had been defections from the ranks during his lifetime by such early favorites as Adler, Jung, and Rank, so that the central core of psychoanalysis was defined by and remained dominated by Freud. During his lifetime a unity within a developing theory could be maintained by accepting his leadership; after his death those who remained in the mainstream of psychoanalysis felt they needed to be free to modify their views, as Freud would have done were he still alive, but they felt some obligation to show that they followed Freud's main teachings in reference to



both method and theory, enough at least to preserve unity among themselves. This inevitably resulted in a certain amount of in-group rigidity and out-group antagonism, understandable in view of the social history of psychoanalysis during Freud's lifetime. Within institutionalized psychoanalysis the main cleavage has come between those who hold to the instinct theory and the universality of the Oedipus complex, on the one hand, and those who believe in the relatively greater importance of sociocultural influences and learning, on the other.

There are some consequences for research that lie in the history of psychoanalysis as a movement. Earlier psychoanalysts, whose motivations were primarily therapeutic and who were untrained in the use of experimental methods in the investigation of psychoanalytic phenomena, tended to disparage the use of such methods. Some still do, believing that the essence of psychoanalysis is displayed only in the continuous relationship between the patient and analyst, resting upon free association in the context of transference. For them, special research on psychoanalysis is often believed to be unnecessary, and simulation of psychoanalysis outside the patient-therapist relationship is thought to be either trivial or misleading. Because of their commitment to the method of psychoanalysis, younger psychoanalysts familiar with research methods seldom turn their attention to psychoanalysis as being itself in need of investigation, but instead study psychotherapy, which is less codified, or turn to topical investigations which do not question the central tenets of psychoanalysis. Those who favor an examination of the role of sociocultural influences are likely to select their problems more as anthropologists or sociologists select theirs. Now that many years have passed since Freud's death, there are signs that these institutional influences upon research are lessening. However, the official institutes for the training of psychoanalysts have as yet given relatively little attention to research; for example, in a careful account of training in psychoanalysis in the United States (Lewin & Ross 1960), only 4 of 460 pages were devoted to research, and it was reported that only 3 of the 14 institutes had brought research at all prominently into the training pattern. The task of consolidating psychoanalysis as a science will continue to call for the collaboration of others if it is to move ahead with any rapidity.

The others who have interested themselves in psychoanalytic research are primarily psychologists, sociologists, and anthropologists, each with his own methods and preoccupations. There was great enthusiasm for investigating psychoanalytic

problems after World War II, and a number of social scientists underwent both personal analyses and training in the psychoanalytic institutes in order to qualify themselves better for such research. However, as pointed out by Shakow (1962), in the early 1960s there was evidence of a growing dissatisfaction and disenchantment with psychoanalysis even on the part of persons fundamentally sympathetic to it. He believed that these reactions were due to the relative absence of research and scholarly activity around psychoanalysis proper, with the result that psychoanalysis was not attracting its share of "the restless souls with questioning minds who were attracted to it in the early days." Shakow recommended that serious consideration be given to the psychoanalytic education of social and behavioral scientists for research.

Psychoanalysis, as a body of facts and theories about behavior and about therapy, cannot have any official spokesmen. The facts and theories are in the public domain, to be studied by any who are interested and competent. Therefore, one interested in psychoanalysis as science need not be concerned about the problems within psychoanalysis as a movement, although it is inevitable that the attitudes of the official psychoanalysts will make the field more or less inviting as an area in which to do research or in which to make a research career.

### Psychoanalytic observation as research

Kris (1947) distinguished between the validation of psychoanalytic propositions through psychoanalytic observation and other "objective methods." He pointed out that the total course of the association between observer and subject provides the equivalent of an experimental setting which permits the testing of long-range predictions; each interview, he said, potentially provides a number of opportunities for the testing of forecasts. The simplest examples of confirmatory evidence are given through interpretations that help the subject to fill in gaps of recall, later confirmed by evidence for the veridicality of the recall. Thus Frenkel-Brunswik (1940) reported an illustration from her own analysis in which the analyst interpreted her reactions as a "Cordelia motive," referring to the youngest daughter of King Lear, who, while the most deserving, lost favor to the other daughters who were more flattering to him. She denied having ever read *King Lear* and at first refused the interpretation, but she later discovered that at the age of 15 she had copied all of Cordelia's lines. Thus we have evidence of repression and support for an analytic reconstruction. Kris recognized

that Freud had used this method successfully on himself, although he saw that there were many issues within psychoanalysis that did not get resolved by psychoanalytic observations and required other methods. Kubie (1953), while recognizing many contributions of psychoanalytic observation to basic science, detailed the following limitations within current psychoanalysis as a scientific research method: limitations with respect to the gathering of primary data, conceptual limitations and ambiguities, lack of quantitative methods of validation, lack of control groups, and deficiencies in specific instruments for measuring many of the processes whose interplay lies at the heart of psychoanalytic theory.

Freud's observations within psychoanalysis (including his own) served as the sole basis for his many insights, and doubtless fresh insights will continue to come from such observations by other psychoanalysts. In view of the limitations noted by both Kris and Kubie, however, it is essential that these observations be supplemented by data of a different kind if psychoanalytic science is to progress.

**Appropriateness of experimental investigation.** A full-fledged psychoanalysis is such a rich and complex process that dissecting out parts of it for special study presents hazards. According to Rapaport, most of those who tried to test psychoanalytic theory either did not understand it or they ignored what they knew: "The overwhelming majority of experiments designed to test psychoanalytic propositions display a blatant lack of interest in the meaning, within the theory of psychoanalysis, of the propositions tested" (1959, p. 142). Although the warning should be heeded that psychoanalysis must be understood if relevant experiments are to be designed, there are ample precedents for gaining insight into psychoanalysis outside the analytic hour itself. While Freud was in no sense an experimentalist, he set the stage for considering observations made outside the analytic hour not only in his *Psychopathology of Everyday Life* but also in his analyses of historical figures, such as Leonardo da Vinci.

It is in the nature of experimental science that it must abstract from the totality of actual events and thus lose some of the richness of events in nature. This applies as inevitably to psychoanalytic investigations as to other kinds of scientific research. The problem confronting the investigator is not that of being made to study everything at once, but of being careful not to simplify his conditions in such a way as to sacrifice relevance to significant aspects of theory. Experimental tests of the

Freudian theory of repression illustrate these problems. Early investigators, doing what Rapaport said others have done, oversimplified the theory by making it assert that anything pleasant is favored in recall over anything unpleasant. Hence they tested whether the words "quinine" and "bitter" would be forgotten more readily than "sugar" and "sweet" from a list containing these and affectively neutral words. Zeller (1951), representative of later experimenters, created a paradigm for inducing repression through experienced failure and then lifted the repression, thus coming much closer to psychoanalytic theory. Others have gone on to show that there are individual differences, so that the circumstances which produce repression for one subject do not do so for others (e.g., Truax 1957). With such refinements, experiments gradually become more sophisticated and relevant.

Not all aspects of psychoanalysis are open to experimental inquiry, but the ingenuity of investigators has opened up areas of study that at first seemed unpromising. Thus it has not proved possible to assert in advance which propositions were most suitable for study. Because psychoanalysis is not an internally consistent and highly integrated theory, the investigator has had to choose which propositions to test or has had to formulate as best he could a few precise propositions for testing. The purpose of experimental study is not only to test or validate the theoretical structure as it now stands, but also to correct and refine it. A scientific psychoanalysis, developed from and supported by controlled investigations, doubtless will differ in some respects from contemporary psychoanalytic theory.

There is by now a sizable literature showing that many workers have been able to formulate testable propositions on the basis of psychoanalytic theory and practice. Freud was raised in the scientific tradition of Hermann von Helmholtz and Ernst Brücke that made him propose psychoanalysis in a form such that much of it is subject to test, at least in principle. That is, it is stated as a deterministic system with lawful relationships expressed in terms of force and energy. While these force and energy concepts have proved troublesome in trying to conceptualize psychoanalysis, they are to be contrasted, for example, with the romanticism and mysticism of Jung. The differences between Freud and Jung are somewhat comparable to the differences between Darwin and Lamarck. Thus Darwin's position, unlike that of Lamarck, permitted the congenial development of the mathematical approach



of modern genetics to the problems of evolution, although Darwin was not himself mathematical. In a somewhat similar sense, Freud, in contrast to Jung, expressed his theories in a form that makes possible experimental testing, although he was not an experimentalist. [See *ANALYTICAL PSYCHOLOGY and the biography of JUNG.*]

### Some representative empirical approaches

In dealing with experimental studies of psychoanalysis, emphasis need not be placed on a narrow definition of experiment; the main point is that some aspects of psychoanalysis are open to empirical studies done in the spirit of experimental science—that is, using methods that are objective, quantitative, and controlled. For example, when apparatus is attached to the patient in order to measure physiological changes associated with verbally expressed anxiety, this approximates an experiment, even though the interaction between the therapist and patient is not experimentally manipulated.

In what follows, instead of organizing experimental studies around an outline of psychoanalytic propositions, nine disparate approaches to the investigation of psychoanalysis—as both a theory of behavior and a therapeutic technique—are used to illustrate the possibilities open for research. Within each of these, one or more specimen studies are cited.

**The study of psychoanalytic psychotherapy.** The recording of what goes on within psychoanalysis by means of electromagnetic tape or sound motion pictures has opened up psychoanalysis for detailed study. It yields more precise and detailed observations than would otherwise be possible, and the data can be used for hypothesis testing, thus approaching the conditions of experimental investigation. One illustration of subsequent analyses of two recorded psychotherapeutic interviews is provided in a symposium in which psycholinguistic aspects of the communication between therapist and patient were studied in relation to some physiological changes that were also recorded (American Psychiatric Association 1961). A number of related studies have not yet come to fruition.

The data obtained by appropriate recording techniques are not only detailed and accurate, but they are also in permanent form and can be reviewed for various purposes. Various content analyses are possible (based, for example, on type-token ratios, silences, kinesic and other paralinguistic features, dyadic interactions, the effects of interpretations). It is possible to place the coded

material, with various transitional probabilities, into a computer and then test hypotheses by asking questions of the computer.

One of the unresolved problems in this kind of study is the degree to which the analysis should be microscopic or macroscopic (thematic). To some extent, each analytic hour has a "topic," and it may be that this will in some cases be lost in the detailed linguistic analysis. The answers will not come a priori but only after much work has been done to find the methods most fully appropriate to the subject matter under investigation. Care has to be taken not to obtain so much data that there is an information overload upon the investigator; selectivity with due regard for objectivity and criteria of relevance will have to be worked out here as in all quantitative investigations.

**Outcome versus process studies.** An important distinction has to be made within studies of psychoanalysis and psychotherapy between *outcome* studies and *process* studies. While there is a place for outcome studies ("What proportion of the treated patients improve or get well?"), these are extremely difficult to do in the case of psychotherapy, for a great many reasons: (1) Diagnosis of the severity of the presenting symptoms can be made on a quantitative basis with only the greatest of difficulty. For one thing, a patient who appears more severely disturbed may be in a temporary crisis that would disappear spontaneously, while another patient who appears less disturbed may have a deep-seated character disorder very resistant to change. A long-standing illness, which is usually more difficult to cure than one of more recent origin, may be brought to the therapist's attention because for some reason the patient is now ready to be cured and presents himself for treatment at a time when the therapeutic result can be quickly achieved. (2) Estimates of degree of improvement are hard to obtain, even with careful follow-ups, for the patient and his relatives are subject to the same difficulties in estimating improvement that the therapist faces in diagnosing severity of ailment. (3) The degree of skill of the therapist and his appropriateness for the given patient are not sufficiently stated by a record of his training or by a knowledge of the general method that he claims to use. (4) Appropriate control groups of patients with similar severity of symptoms treated either by other methods, or not treated at all, are hard to establish. The reported success of easier and superficially more objective therapies, such as the so-called behavior therapies, have to be judged against the successes of all sorts

of magical and faith cures that take place without the benefit of science. Although outcome studies are difficult, improved ones are very much needed to establish the therapeutic value of alternative therapies. [See MENTAL DISORDERS, TREATMENT OF, *article on* BEHAVIOR THERAPY.]

Studies of *process*, that is, of movement (whether progressive or regressive) taking place within and between psychotherapeutic sessions, are at present much more promising, although these too are hard to design and the results yielded to date are not very striking. In the careful study by Bellak and Smith (1956) of progress in therapy, it was found that the judges were in substantial agreement about the patients, but what they agreed on was the persisting personality structure of the patients, not the dynamic changes taking place as therapy progressed. [See MENTAL DISORDERS, TREATMENT OF, *article on* PSYCHOLOGICAL TREATMENT.]

**Miniaturized procedures with nonpatients.** Rather than study the records of a patient in psychoanalysis, it is possible to select out certain features, such as free association, and study them. A start in this direction was made by Colby (1960a), who had a subject free associate alone with a tape recorder and then with a silent observer present; the differences in the subject's productions were noted. A number of possible parameters come to mind: sitting up, lying down, facing or not facing the therapist, note taking and recording, the number of days between sessions, having parallel sessions of husband and wife, mother and child. Many of these focal points recur in treatises on psychoanalytic method, but few have been put to actual comparative tests.

The early studies of Masserman (Masserman & Yum 1946) show how aspects of therapy can be studied through the treatment of experimentally induced neuroses in animals, particularly aspects of "working through" in the overcoming of fears, and aspects of the influence of drugs (alcohol) in relation to conflict. While these studies are limited when done with nonverbal subjects, they pose some searching questions for psychoanalysis. For example, if the experimental neurosis can indeed be defined as a neurosis, it raises doubts about a very central psychoanalytic dictum, namely, that the Oedipus complex constitutes the nuclear complex of every neurosis.

**Hypnosis as a means of study.** The fact that psychoanalysis grew out of Freud's experience with hypnosis, of which the psychoanalytic couch is a historical residual, makes the bridges between

what happens in hypnosis and what happens in psychoanalysis easy to construct. Although Freud turned his back upon hypnosis early in his career, other psychoanalysts have continued to see its relevance. Two illustrations may serve to show some of the many experimental possibilities.

Some psychoanalysts, particularly Franz Alexander, have proposed what has been called the specificity hypothesis regarding psychosomatic symptoms—that is, that under defined kinds of stress some people will give symptoms of one kind while others will yield different symptoms. Using hypnosis, Graham, Stern, and Winokur (1958) induced differential attitudes hypothesized as appropriate to Raynaud's disease and to urticaria (hives) in subjects not suffering these ailments. When these subjects were placed under stress, the physiological responses were indeed found to be the ones more nearly appropriate to the ailment corresponding to the induced attitude.

In somewhat related experiments, Kehoe and Ironside (1963) produced depressive reactions under hypnosis and then studied both the gastric secretions and other correlates of the experimentally evoked depression.

The possible services of hypnosis in the study of various motivational and psychodynamic problems have been reviewed by Hilgard [1964; *see also* HYPNOSIS].

**Psychophysiological studies.** In establishing a functional system, Freud somewhat reluctantly turned away from the neurophysiology in which he was trained, which he had used extensively and not unwisely in his early theorizing. At a number of times, however, Freud indicated that he hoped eventually for a physiological basis for psychoanalysis (e.g., 1920). Some contemporary investigators in the field of psychoanalysis believe that the time is now ripe to provide a better neurophysiological anchorage for psychoanalytic propositions, and their research endeavors illustrate this concern. Weiner and others (1957) were able to predict which of a number of inductees would become peptic ulcer casualties as a result of the strains of military service. They used data on serum pepsinogens in defining the sample, and then, using psychological information alone, predicted that ten cases among a much larger number would be most likely to develop ulcers during basic training. Of these, seven had proven ulcers before the end of training; of the remaining inductees in the sample, only two developed ulcers. Hence, the psychological tests not only had picked out the promising cases but had missed very few.



The psychological factors emphasized as predisposing to ulcers were an intense need to maintain relations with others, a need to please and placate authority figures, and fears of expressing hostility or of experiencing loss.

The newer physiological methods of studying dreams have made possible another contact between experiment and psychoanalysis. Thus the function of dreams as tension reducers is supported by studies of dream deprivation in which the sleeper "catches up" by dreaming more frequently on nights following those during which his dreaming has been interrupted [Dement 1960; *see also DREAMS and PSYCHOSOMATIC ILLNESS*].

**Studies of motivational conflicts.** Psychoanalysis can be considered a motivational psychology with primary emphasis upon motivational conflicts and their consequences. These areas lend themselves well to experimental study, including animal studies. One of the experimental studies of the derivatives of conflicting motives is that of Clark (1952), in which he aroused sexual motives by displaying nude pictures to college students and studied the overt and symbolic expression of sex under various conditions, including mild intoxication at a fraternity beer party. He found that anxiety under ordinary circumstances (for example, being in a college classroom) inhibited manifest sexual expression but gave rise to increased symbolic expression of sex; the reverse occurred under the influence of alcohol—that is, there was more manifest sexual expression and less symbolic expression [*see CONFLICT, article on PSYCHOLOGICAL ASPECTS; MOTIVATION*].

**Studies of cognitive processes.** There is by now a large literature based upon tachistoscopic perception and related methods, much of it motivated by an interest in psychoanalytic concepts. Many of the studies on tachistoscopic perception relevant to psychoanalysis were reviewed by Charles Fisher in the introduction to a monograph translating two classical studies (Pötzl et al. 1960).

We may take as representative of these the study by Paul and Fisher (1959) which investigated the hypothesis that the exposure of a figured visual stimulus below the level of awareness would produce images that would reflect, both directly and in transformed ways, the influence of the stimulus, its meanings, and its forms. The experimental procedure was to expose a simple double-profile figure within a circular outline at 1/100 of a second, slightly out of focus; this alternated with the exposure of a blank slide. After each exposure the subject was asked to draw whatever images

came to mind. These drawings were then sorted by a number of experienced judges as more likely to have followed the figured stimulus or the blank one. The judges were only moderately successful, but taken together the results were highly significant statistically, indicating some influence from the exposed figure. From the published figures it is evident that there was little direct representation of the exact figure as exposed; the most directly representative figures consisted of overlapping or multiple circles. Some degree of transformation would, of course, be implied by psychoanalytic theory as well as by other theories [*see ATTENTION and PERCEPTION, article on UNCONSCIOUS PERCEPTION*].

**Studies of learning and retention.** The concept of motivated forgetting is represented in psychoanalysis through some aspects of repression, in which memories, while active, are not available to direct recall.

Cleaves (1964) gave a Kent-Rosanoff test to subjects of known hypnotic susceptibility in order to find the words which in the normal waking state were affect-laden and caused difficulties in response ("complex indicators"), such as long reaction times. He hypothesized that these words should be appropriate targets for repression relative to words more neutral for the subject. He tested the hypothesis by constructing for each subject a list of words to be memorized, half neutral and half affect-laden for him. The list was memorized under hypnosis, and then the suggestion was given that half the words would be forgotten (post-hypnotic amnesia) until a release signal was given. Counting only the words that fitted the paradigm of repression (temporarily forgotten but recovered), he found that these were disproportionately the target words selected on the basis of the word association test. Thus hypnosis served conveniently to test an appropriate hypothesis about repression.

A study of associative priming by Schwartz and Rouse (1961) illustrates another convergence between research formulated within a psychoanalytic frame of reference and the kinds of study generally familiar within the laboratories of academic psychology. These investigators found, for example, that in a test of immediate memory words that were forgotten had nevertheless "primed" their associates, so that the associated words were selected beyond chance expectation on a subsequent recognition test from which the original stimulus words were omitted. Thus their results bear on free association in relation to recall, as

well as upon the experimental psychology of learning, recall, and recognition.

**Maturation and development.** Psychoanalysis is also a developmental psychology. The formal theory of psychosexual stages is well known, with the child experiencing and coping with various stages of development on the way to maturity (the characteristics of each of which have been correlated under the general headings termed oral, anal, phallic, latency, and genital stages). The addition of the psychosocial stages by Erikson (1959) has enriched the theory, showing that eventually early identifications become welded into either a healthy or a pathological identity. He also carried the scheme through adolescence into later maturity. Erikson's psychosocial scheme is an appealing one, but so far it has received essentially literary treatment; it needs both theoretical and empirical clarification [see **IDENTITY, PSYCHOSOCIAL**].

The success with which Whiting and Child (1953) were able to predict, on the basis of reported child-rearing practices, certain beliefs about illness on the part of adults in many nonliterate cultures gives some support to the significance of early childhood in shaping personality structure. Of course, these same children are brought up by the very adults who trained them in childhood, so that the impact of their training is not over once and for all. A learning-theory interpretation fits the facts as well as a psychoanalytic one, and indeed, Whiting and Child use such an interpretation. Yet the influences they chose to study (feeding, toilet training, sex, aggression, dependency) have their origins in psychoanalytic modes of thought. The corresponding animal studies, initiated by Hunt (1941), showed that white rats which experienced early feeding frustration hoarded food excessively later on, when new frustrations presumably reactivated the early patterns of insecurity.

The attempts to support psychoanalytic conceptions about the importance of breast feeding, weaning, toilet training, and the like have led to somewhat equivocal results, with Orlansky (1949) concluding from his review of cross-cultural data that the ground plan of formal psychosexual stages through which infantile libidinal drives are channeled is less satisfactory than some sort of sociocultural interpretation. However, it is clear that the answers have not all been found, and psychoanalysis has made an abiding contribution by calling attention to the significance of the early years and of appropriate mothering to personality

development [see **PERSONALITY, article on PERSONALITY DEVELOPMENT; SOCIALIZATION**].

**Personality structure, defenses, and cognitive styles.** Psychoanalytic personality theory assumes that in the course of his history (particularly, but not exclusively, early history), the individual achieves a personality structure which is relatively stable. If some kinds of influences are extreme, a personality "syndrome" such as anal compulsivity may emerge, with the triad of traits—orderliness, stinginess, and obstinacy (Sears 1943). Another such syndrome is said to be the authoritarian personality (Adorno et al. 1950). These personality structures are relatively stable, but may change slowly, as within a successful psychoanalysis. [See **PERSONALITY, POLITICAL, article on CONSERVATISM AND RADICALISM**.]

Empirical assessment of personality structures, cognitive styles, defense preferences, and coping mechanisms has assumed large proportions, one evidence of it being the prominence of projective tests (Rorschach inkblots, the Thematic Apperception Test, Blacky dog pictures) designed to reveal aspects of the personality structure not always manifested in outward behavior [see **DEFENSE MECHANISMS**].

The literature on cognitive styles is represented by Witkin and others (1962) and by Gardner and others (1959). Witkin and his associates have been concerned with field dependency and field independency, aspects of cognition with personality correlates. Gardner and his associates have given more emphasis to such distinctions as that between the sensitizers and the levelers. While these studies have a background in gestalt conceptions as well as in psychoanalytic ones, what makes them relevant to psychoanalysis is their demonstration that personality structure and momentary needs affect cognitive activity. This is, of course, a prominent conception in psychoanalysis—as, for example, in rationalization and projection.

Studies of a more ecological character are also possible, as in the investigations of differences in defense preferences of children of blue-collar workers and white-collar workers (Miller & Swanson 1960). It is difficult to intervene experimentally in the upbringing of children to study the long-term effects, but it is possible to capitalize upon some naturally occurring events which produce similar data to those that would be obtained from a controlled experiment. One such naturally occurring event is the loss in childhood of a parent through death. The consequence of such a loss can be studied in groups as though an experiment had



been undertaken to see what happens to the child if a parent is permanently removed. Such studies have been done and through a combination of statistical and clinical observations have reached two major conclusions. First, the age at which a mother has been lost in childhood provides a target date for personality disturbance or mental illness, the illness being triggered when the child who suffered loss has grown to adulthood and has a child of the age at which the loss occurred (J. R. Hilgard & Newman 1959). Second, the incidences of certain mental illnesses, particularly schizophrenia, are related to these losses of a parent in childhood (J. R. Hilgard & Newman 1961). Such studies give support to the role of parents in normal personality development and to the reactivation at times of adult crisis of the stresses associated with periods of deprivation in childhood.

Because too many inferences are involved it is not possible to draw up a scoreboard to mark the successes and failures of experiments in validating psychoanalytic propositions. However, those non-psychoanalysts who have summarized experimental investigations of psychoanalytic theories have found a great deal of the evidence favorable to psychoanalysis (Blum 1953; Farrell 1951; Hilgard 1952; Hilgard & Bower 1966; Sarnoff 1962; Sears 1943). Unfortunately they have not been able to point to many experimental findings that have clarified and strengthened psychoanalysis, or, indeed, that have had much impact on either its theories or its practices. The full impact of experiment will not be felt until psychoanalysis is moved forward and enriched by the results that are reported. Experimental studies cannot be expected to do all the work of theoretical reconstruction as psychoanalysis advances and becomes assimilated within social science, but the interaction between experiment and theory should prove to be a vital part of that reconstruction.

ERNEST R. HILGARD

[Directly related are the entries EXPERIMENTAL DESIGN; MENTAL DISORDERS, article on EXPERIMENTAL STUDY. Other relevant material may be found in FANTASY; INTERVIEWING, article on THERAPEUTIC INTERVIEWING; OBSERVATION; PROJECTIVE METHODS.]

#### BIBLIOGRAPHY

- ADORNO, THEODOR W. et al. 1950 *The Authoritarian Personality*. American Jewish Committee, Social Studies Series, No. 3. New York: Harper.
- AMERICAN PSYCHIATRIC ASSOCIATION 1961 *Comparative Psycholinguistic Analysis of Two Psychotherapeutic Interviews*. New York: International Universities Press.
- BELLAK, LEOPOLD 1961 Research in Psychoanalysis. *Psychoanalytic Quarterly* 30:519-548.
- BELLAK, LEOPOLD; and SMITH, M. BREWSTER 1956 An Experimental Exploration of the Psychoanalytic Process. *Psychoanalytic Quarterly* 25:385-414.
- BLUM, GERALD S. 1953 *Psychoanalytic Theories of Personality*. New York: McGraw-Hill.
- CLARK, RUSSELL A. 1952 The Projective Measurement of Experimentally Induced Levels of Sexual Motivation. *Journal of Experimental Psychology* 44:391-399.
- CLEMES, STANLEY C. 1964 Repression and Hypnotic Amnesia. *Journal of Abnormal and Social Psychology* 69:62-69.
- COLBY, KENNETH M. 1960a Experiment on the Effects of an Observer's Presence on the Imago System During Psychoanalytic Free-association. *Behavioral Science* 5:216-232.
- COLBY, KENNETH M. 1960b *An Introduction to Psychoanalytic Research*. New York: Basic Books.
- CONFERENCE ON EXPERIMENTAL PSYCHIATRY, WESTERN PSYCHIATRIC INSTITUTE AND CLINIC, 1959 1961 *Lectures on Experimental Psychiatry*. Univ. of Pittsburgh Press.
- CONFERENCE ON RESEARCH IN PSYCHOTHERAPY, 1958 1959 *Research in Psychotherapy: Proceedings*. Edited by Eli A. Rubinstein and M. B. Parloff. Washington: American Psychological Association.
- CONFERENCE ON RESEARCH IN PSYCHOTHERAPY, 1961 1962 *Research in Psychotherapy: Proceedings*. Edited by Hans H. Strupp and L. Luborsky. Washington: American Psychological Association.
- DEMENT, WILLIAM 1960 The Effect of Dream Deprivation. *Science* 131:1705-1707.
- ERIKSON, ERIK H. 1959 Identity and the Life Cycle: Selected Papers. *Psychological Issues* 1, Monograph no. 1.
- FARRELL, BRIAN A. (1951) 1954 The Scientific Testing of Psychoanalytic Findings and Theory. Pages 448-457 in Howard Brand (editor), *The Study of Personality*. New York: Wiley. → First published in Volume 24 of the *British Journal of Medical Psychology*.
- FEIGL, HERBERT; and SCRIVEN, MICHAEL (editors) 1956 *The Foundations of Science and the Concepts of Psychology and Psychoanalysis*. Minnesota Studies in the Philosophy of Science, Vol. 1. Minneapolis: Univ. of Minnesota Press.
- FRENKEL-BRUNSWIK, ELSE 1940 Psychoanalysis and Personality Research. *Journal of Abnormal and Social Psychology* 35:176-197.
- FRENKEL-BRUNSWIK, ELSE 1954 Psychoanalysis and the Unity of Science. *Dædalus: Journal of the American Academy of Arts and Sciences* 80:271-350.
- FREUD, SIGMUND (1920) 1950 *Beyond the Pleasure Principle*. Authorized translation from the 2d ed., by C. J. M. Hubback. International Psychoanalytic Library, No. 4. New York: Liveright. → First published under the title *Jenseits des Lustprinzips*. A paperback edition, translated by James Strachey, was published in 1959 by Bantam Books.
- GARDNER, RILEY W. et al. 1959 Cognitive Control: A Study of Individual Consistencies in Cognitive Behavior. *Psychological Issues* 1, Monograph no. 4.
- GILL, MERTON M. 1963 Topography and Systems in Psychoanalytic Theory. *Psychological Issues* 3, Monograph no. 10.
- GRAHAM, DAVID T.; STERN, JOHN A.; and WINOKUR, GEORGE 1958 Experimental Investigation of the Specificity of

- Attitude Hypothesis in Psychosomatic Disease. *Psychosomatic Medicine* 20:446-457.
- HALL, CALVIN S.; and LINDZEY, GARDNER 1954 Psychoanalytic Theory and Its Applications in the Social Sciences. Volume 1, pages 143-180 in Gardner Lindzey (editor), *Handbook of Social Psychology*. Cambridge, Mass.: Addison-Wesley.
- HILGARD, ERNEST R. (1952) 1956 Experimental Approaches to Psychoanalysis. Pages 3-45 in Eugene Pumphlin-Mindlin (editor), *Psychoanalysis as Science*. New York: Basic Books.
- HILGARD, ERNEST R. (1962) 1965 The Scientific Status of Psychoanalysis. Pages 184-205 in Irwin G. Sarason (editor), *Science and Theory in Psychoanalysis*. Princeton, N.J.: Van Nostrand.
- HILGARD, ERNEST R. 1964 The Motivational Relevance of Hypnosis. Volume 12, pages 1-44 in *Nebraska Symposium on Motivation*. Edited by David Levine. Lincoln: Univ. of Nebraska Press.
- HILGARD, ERNEST R.; and BOWER, GORDON H. 1966 Freud's Psychodynamics. Pages 264-296 in Ernest R. Hilgard, *Theories of Learning*. 3d ed. New York: Appleton. → First published in the second (1958) edition of Hilgard's *Theories of Learning*.
- HILGARD, JOSEPHINE R.; and NEWMAN, MARTHA F. 1959 Anniversaries in Mental Illness. *Psychiatry* 22:113-121.
- HILGARD, JOSEPHINE R.; and NEWMAN, MARTHA F. 1961 Evidence for Functional Genesis in Mental Illness: Schizophrenia, Depressive Psychoses and Psychoneuroses. *Journal of Nervous and Mental Disease* 132: 3-16.
- HOOK, SIDNEY (editor) 1959 *Psychoanalysis, Scientific Method and Philosophy*. New York Univ. Press. → A paperback edition was published in 1960 by Grove.
- HUNT, JOSEPH McV. 1941 The Effects of Infant Feeding-frustration Upon Adult Hoarding in the Albino Rat. *Journal of Abnormal and Social Psychology* 36: 338-360.
- JANIS, IRVING L. 1958 The Psychoanalytic Interview as an Observational Method. Pages 149-181 in Gardner Lindzey (editor), *Assessment of Human Motives*. New York: Rinehart.
- KEHOE, MICHAEL; and IRONSIDE, WALLACE 1963 Studies on the Experimental Evocation of Depressive Responses Using Hypnosis. II: The Influence of Depressive Responses Upon the Secretion of Gastric Acid. *Psychosomatic Medicine* 25:403-419.
- KRIS, ERNST 1947 The Nature of Psychoanalytic Propositions and Their Validation. Pages 239-259 in Sidney Hook and Milton R. Konvitz (editors), *Freedom and Experience*. A publication of the New School for Social Research. Ithaca, N.Y.: Cornell Univ. Press.
- KUBIE, LAWRENCE S. 1953 Psychoanalysis as a Basic Science. Pages 120-145 in Franz Alexander and Helen Ross (editors), *20 Years of Psychoanalysis*. New York: Norton.
- LEWIN, BERTRAM D.; and ROSS, HELEN 1960 *Psychoanalytic Education in the United States*. New York: Norton.
- MASSERMAN, JULES H.; and YUM, K. S. 1946 An Analysis of the Influence of Alcohol on Experimental Neuroses in Cats. *Psychosomatic Medicine* 8:36-52.
- MILLER, DANIEL R.; and SWANSON, GUY E. 1960 *Inner Conflict and Defense*. New York: Holt.
- ORLANSKY, HAROLD 1949 Infant Care and Personality. *Psychological Bulletin* 46:1-48.
- PAUL, J. H.; and FISHER, CHARLES 1959 Subliminal Visual Stimulation: A Study of Its Influence on Subsequent Images and Dreams. *Journal of Nervous and Mental Disease* 129:315-340.
- PÖTZL, OTTO; ALLERS, RUDOLF; and TELER, JAKOB 1960 Preconscious Stimulation in Dreams, Associations, and Images: Classical Studies. *Psychological Issues* 2, Monograph no. 7.
- PUMPHLIN-MINDLIN, EUGENE (editor) (1952) 1956 *Psychoanalysis as Science*. New York: Basic Books.
- RAPAPORT, DAVID 1959 The Structure of Psychoanalytic Theory: A Systematizing Attempt. Pages 55-183 in S. Koch (editor), *Psychology: A Study of a Science*. Volume 3: Formulations of the Person and the Social Context. New York: McGraw-Hill.
- SARASON, IRWIN G. (editor) 1965 *Science and Theory in Psychoanalysis*. Princeton, N.J.: Van Nostrand.
- SCHWARTZ, FRED; and ROUSE, RICHARD O. 1961 The Activation and Recovery of Associations. *Psychological Issues* 3, Monograph no. 9.
- SEARS, ROBERT R. 1943 *Survey of Objective Studies of Psychoanalytic Concepts*. Bulletin No. 51. New York: Social Science Research Council.
- SHAKOW, DAVID 1962 Psychoanalytic Education of Behavioral and Social Scientists for Research. Pages 146-161 in Academy of Psychoanalysis, *Science and Psychoanalysis*. Volume 5: Psychoanalytic Education. New York: Basic Books.
- SHAKOW, DAVID; and RAPAPORT, DAVID 1964 The Influence of Freud on American Psychology. *Psychological Issues* 4, Monograph no. 13.
- Symposium on Psychoanalysis and Validation. 1961 *British Journal of Medical Psychology* 34:1-21.
- TRUAX, CHARLES B. 1957 The Repression Response to Implied Failure as a Function of the Hysteria-Psychasthenia Index. *Journal of Abnormal and Social Psychology* 55:188-193.
- WEINER, HERBERT et al. 1957 Etiology of Duodenal Ulcer. I: Relation of Specific Psychological Characteristics to Rate of Gastric Secretion (Serum Pepsinogen). *Psychosomatic Medicine* 19:1-10.
- WHITE, ROBERT B. 1961 Recent Developments in Psychoanalytic Research. *Diseases of the Nervous System* 22:609-616, 671-679.
- WHITING, JOHN W. M.; and CHILD, IRVIN L. 1953 *Child Training and Personality: A Cross-cultural Study*. New Haven: Yale Univ. Press.
- WITKIN, HERMAN A. et al. 1962 *Psychological Differentiation: Studies of Development*. New York: Wiley.
- ZELLER, ANCHARD F. 1951 An Experimental Analogue of Repression. III: The Effect of Induced Failure and Success on Memory Measured by Recall. *Journal of Experimental Psychology* 42:32-38.

## PSYCHOGENETICS

See GENETICS, article on GENETICS AND BEHAVIOR.

## PSYCHOLINGUISTICS

See LANGUAGE, article on THE PSYCHOLOGY OF LANGUAGE; LINGUISTICS, article on THE FIELD.

## PSYCHOLOGICAL ANTHROPOLOGY

See CULTURE AND PERSONALITY.



## PSYCHOLOGICAL SYSTEMS

See under SYSTEMS ANALYSIS.

## PSYCHOLOGICAL WARFARE

*Psychological warfare* is a term that came into vogue in the United States during World War II to identify an activity as old as the history of conflict. Although the term gained wide currency in popular and scientific discussions in the United States and Europe within a decade after the end of World War II, meanings ascribed to it were not always clear, and there were several views concerning its nature and scope. Some have limited its meaning to that range of activity which clearly falls within the jurisdiction of the armed forces and which is centered on the dissemination of propaganda to specified target audiences for the purpose of supporting the attainment of a given military mission. This view of psychological warfare, however, is too limited in scope for most journalists, politicians, and scholars who discuss the subject.

As used in popular and scientific journals, a much wider variety of meanings has been attached to the term. First, there is the view that psychological warfare is the sum total of a nation's efforts to influence the opinions and behavior of foreign peoples and governments in desired directions through means other than the employment of a nation's political, economic, and military resources. Those who subscribe to this view generally agree that propaganda is the principal, but not the sole, ingredient of psychological warfare.

Second, there is the view that psychological warfare involves an even wider range of activity, including symbolic acts of violence and terror designed to intimidate or to persuade an adversary to adjust his behavior. Those who hold to this view of psychological warfare would include within its scope various undercover activities such as espionage and subversion, assassinations and other forms of terrorism, and censorship, when they are designed to mold the opinion or behavior of specific groups.

A third view holds that the term includes such activities as the premeditated twisting of personality through techniques popularly described as "brainwashing," hypnosis, and the employment of psychopharmaceutical agents allegedly used by some communist states [see BRAINWASHING].

There is still a fourth view, which holds that psychological warfare includes the molding of public attitudes of one's own people and extends across the spectrum of political-military action to hit-and-

run guerrilla raids and other acts of a paramilitary character conducted in enemy rear areas. Thus, at times the term has been used as though it were synonymous with political indoctrination or orientation, political persuasion, indirect aggression, "protracted conflict," or a "strategy of subversion."

In the twenty years that followed the end of World War II, doctrine concerning psychological warfare in the United States and western Europe—its nature, its uses, and its role in modern international relations—tended to coalesce to form a consensus. Thus, increasingly, those who discuss it accept the basic conclusion that unless a technique involves the premeditated manipulation of opinion through the use of one or more of the media of communications it does not involve psychological warfare.

## History

The British military analyst and historian J. F. C. Fuller is believed to have been the first to employ the term "psychological warfare"—in 1920—although the activities it describes go back to ancient times. In discussing implications to be drawn from World War I advances in military technology, he suggested that traditional means of warfare may in time be "replaced by a purely psychological warfare, wherein weapons are not even used or battlefields sought . . . but [rather] . . . the corruption of the human reason, the dimming of the human intellect, and the disintegration of the moral and spiritual life of one nation by the influence of the will of another is accomplished" (1920, p. 320). Fuller's use of the term at the time attracted no unusual attention in either British or American military and scholarly circles. Twenty years later, as the British were "tooling up" for a propaganda effort similar to the one that General Fuller had forecast, they adopted the expression "political warfare." Only in January 1940 did the term come into American use, when an article was published entitled "Psychological Warfare and How to Wage It." Over a year later the Committee for National Morale used the term in the title of a book (Farago & Gittler 1942) but ascribed to it a meaning far different from that in postwar usage.

During World War II "psychological warfare" was widely used to denote the organization and, more especially, the activity of propagandists who served with or supported the U.S. armed forces. But the term was not generally applied to broader, longer-range propaganda efforts. "Information," "strategic services," and "political warfare" were preferred and were used in titles and names of agencies, e.g., Coordinator of Information (COI), Ministry of In-

formation, Office of Strategic Services (OSS), etc. The major British agency for the overseas effort that was later described as psychological warfare was the Political Warfare Executive (PWE). In the United States the civilian-directed effort was entrusted to the Office of War Information (OWI).

The psychological warfare support provided U.S. and allied combat forces during World War II, while significant, was largely the result of *ad hoc* improvisations. Neither the Americans nor the British had made plans before the war to employ combat propaganda. Civilian propagandists from the OWI and PWE were casually included among Eisenhower's forces and organized into propaganda units and staff sections. Later on, experience gained in Africa induced General Eisenhower to create a special staff section for psychological warfare at his headquarters in Europe. The British and American personnel from both civilian and military agencies were selected, trained, and assigned to army group, field army, and army corps headquarters. Psychological warfare operations in the action against Japan were even more improvised: the OWI sent small civilian staffs to major rear-area headquarters—Hawaii, Australia, China, and India. In every instance they first had to convince the theater commander that psychological warfare could make a contribution to victory; next they had to induce military commands to assign qualified personnel to *ad hoc* units; and then it was necessary to win the cooperation of combat commanders to enable these units to practice their skills. Psychological warfare units were generally attached to intelligence sections. Propaganda operations were therefore seldom adequately integrated into operational plans and were generally employed in addition to, not as a part of, regular military operations. The War Department in Washington assigned headquarters personnel to monitor overseas operations only intermittently, and even then no operational or planning function was performed for the overseas commands. The impetus and planning for the use of psychological warfare in overseas areas came from those within each command; the pattern of organization, therefore, varied from one theater of operations to another.

After the war all psychological warfare units and special staff sections were disbanded. The Office of War Information was dissolved, and a greatly curtailed foreign information service was established in September 1945 by executive order as an interim operation under State Department auspices. In 1948 Congress first authorized its continuance. Thus the civilian-sponsored program remained under direct State Department jurisdiction until 1953,

when the U.S. Information Agency (USIA) was established as a quasi-autonomous organization.

During the years that the foreign information program remained under State Department jurisdiction, it increasingly became the practice for the general public to describe its operations as psychological warfare. With the increased tensions growing out of the "cold war," the outbreak of conflict in Korea, and the launching of a "great new campaign of truth," by President Truman in 1950, there was recognition of a need for a new instrument to coordinate the overseas propaganda effort of the several departments and agencies involved. It was natural, in view of the growing acceptance of the term psychological warfare, for a term like psychological strategy to be adopted to label the agency given responsibility for the coordination of the over-all foreign information program. In August 1950 President Truman established a national Psychological Strategy Board (PSB) to coordinate "foreign information and psychological strategy" where more than one agency was involved. In 1953 the PSB was replaced by an Operations Coordinating Board, which in turn was abolished by President Kennedy in 1961.

#### Modern doctrine and methods

Relatively early in the Korean War the U.S. Army created, as a special staff section, the Office of the Chief of Psychological Warfare (OCPW) and a center and school in North Carolina for the codification of doctrine and the training of military personnel in what came to be called, in short, "paywar." With the establishment of OCPW and a paywar center and school, interest in psychological warfare spread throughout the U.S. armed forces. With increased interest came new concepts, definitions, and a crystallization of doctrine. Psychological warfare, as previously defined, gave way to new terms—psychological operations (payops) and psychological activities. Payops were defined as political, military, economic, and ideological actions planned and conducted to create in enemy, hostile, neutral, or friendly foreign groups emotions, attitudes, or behavior favorable to the accomplishment of national objectives.

The American approach to psychological warfare that grew out of World War II and the cold war experience has had a significant impact on the thought and behavior of allied nations. The American term psychological warfare has largely displaced terms these nations previously used. More important than this, however, is the widespread recognition now given to the need for specially trained and equipped units to conduct psycholog-



ical warfare operations in time of crisis. In preparing for this responsibility, many nations have sent officer personnel to the United States for specialized training in American military schools.

Modern psychological warfare on both the national level and within the armed services differs markedly from propaganda operations conducted in the past during times of national crisis in several ways. First, there is official recognition that the psychological warfare element, taken in its broadest sense, is one of four major forces employed during both peace and war to give maximum support to policies in order to increase the probabilities and favorable consequences of victory and lessen chances of defeat. Thus, in the highest councils of government the psychological factor in international relations is considered along with political, economic, and military factors in making major foreign policy decisions.

Second, technological advances in the science of communications have made available to the modern propagandist a variety of means for reaching audiences that previously could only have been reached after great effort and much uncertainty. Short-wave radio transmitters and receivers, films, the growing distribution of television outlets, and the availability of high-speed printing presses combine to increase the likelihood that any given audience a propagandist desires to reach can be exposed to the message he wants it to receive.

Third, modern developments in public opinion polling, audience sampling, panel interviewing, intelligence analysis, and the new techniques for assessing cultural traits of foreign groups enable one to make more accurate predictions of group and mass behavior. Thus it is possible for modern psychological warfare operations to be more nearly manageable and predictable than were the hit-or-miss operations of the past.

**Communications media.** The choice of communications media for psychological warfare operations depends on the target audience to be addressed and the time one has to prepare and deliver the message. In combat situations, printed leaflets and newsheets disseminated by airdrops and voice broadcasts amplified by electronic loudspeakers are the most commonly employed media. Operations conducted from areas farther to the rear or to audiences far removed from combat zones utilize mobile and fixed long-wave and medium-wave radio transmitters.

Media employed in strategic or national-level operations are commonly classified as either slow or fast. Slow media include news magazines, books, pamphlets, motion picture films, and lectures.

They are so described because there is a considerable time lag between the sending and the receiving of the message. Fast media are those that rely largely on electronic communications. In addition to short-wave radio broadcasts, fast media include carefully timed statements by statesmen deemed to be of sufficient interest to be picked up and transmitted as news and a daily news file transmitted by wireless to overseas outlets for local dissemination.

**Intelligence requirements.** Intelligence requirements for effective psychological warfare operations are both enormous and varied. There are three major types of intelligence required. The first is background data, in great detail, concerning the predispositions and vulnerabilities of the target or targets to be addressed. This type of intelligence is described as target analysis. A second type of intelligence needed is that employed in propaganda output. The writings, press releases, and speeches of leading adversaries are thus combed for material that can be usefully employed against the target. Finally, it is necessary to provide a means for checking the results of one's work. Is the message getting through? Is it clearly understood? In what ways can a more favorable response be elicited? Answers to these and other questions are sought by such means as are appropriate to the situation. In time of conflict, and with respect to closed societies, it is sometimes necessary to employ clandestine means for gathering data to assess effectiveness. With respect to societies more or less open to direct approach, variations of techniques employed in domestic public opinion polling and market surveys may be employed.

In both World War II and the Korean War, carefully selected panels of prisoners of war were found to be useful in pretesting and post-testing messages and in making critiques of the format of presentation or the media of communications. Where such panels were used, extreme care had to be taken to insure that they were as nearly representative as possible of target groups yet to be addressed. Furthermore, the fact that prisoners had either defected or had been captured set them apart from those still within enemy ranks, and great care had to be taken in extrapolating findings based on such groups [see INTELLIGENCE, POLITICAL AND MILITARY].

**Research.** Recognition of the need for psychological warfare target intelligence in both depth and breadth has had a great impact on American social science scholarship in the period since the outbreak of the Korean War. Social scientists in the United States, since 1950, have devoted consider-

ably more attention to studies of elites, would-be elites, and communications and other behavioral patterns in foreign societies than in any previous period. Great attention has been given to the identification of research criteria of significance in the study of alien cultures. Likewise, great effort has been devoted to the systematic development of more sophisticated tools for the effective identification and assessment of such material as is useful to psychological warfare practitioners. [See COMMUNICATION, MASS.]

Since 1950 the national government, through the facilities of the State Department, the Agency for International Development (AID), and its predecessors, the USIA and the three service branches of the Defense Department, has subsidized university groups, individuals, and nonprofit organizations to undertake specific research projects that might not otherwise have been undertaken, or tailored them to meet the needs of the so-called "psychological warfare community."

From the findings of this government-sponsored and other research there has emerged a clearer understanding of the requirements for an effective psychological warfare effort in times of peace and of crisis. Whatever media of communications are employed, psychological warfare can contribute to the attainment of a nation's objectives only if the messages transmitted are credible, clearly understood, and seek to elicit a response within the capability of the target audience.

What is credible is not to be equated necessarily with truth. What is credible to any given audience is what it believes to be true, not what is in fact true. Credibility is a complex goal and not easily achieved. What is credible is determined by the audience, not the content of a message. The development of credibility thus must be constantly and persistently pursued in every psychological warfare campaign. Without this important element much of a propagandist's effort will go for naught.

WILLIAM E. DAUGHERTY

[See also COMMUNICATION, POLITICAL; PROPAGANDA; WAR.]

#### BIBLIOGRAPHY

- BARGHOORN, FREDERICK C. 1964 *Soviet Foreign Propaganda*. Princeton Univ. Press.
- BARRETT, EDWARD W. 1953 *Truth Is Our Weapon*. New York: Funk & Wagnalls.
- BUREAU OF SOCIAL SCIENCE RESEARCH, WASHINGTON, D.C. 1956 *International Communication and Political Opinion: A Guide to the Literature*, by Bruce L. Smith and Chitra M. Smith. Princeton Univ. Press.
- CLEWS, JOHN C. 1964 *Communist Propaganda Techniques*. New York: Praeger.

- DAUGHERTY, WILLIAM E. (compiler) 1958 *A Psychological Warfare Casebook*. Baltimore: Johns Hopkins Press. → A compilation of articles and case studies covering various problems, techniques, and requirements of psychological warfare. Contains an extensive list of references.
- DAVIDSON, PHILIP 1941 *Propaganda and the American Revolution: 1763-1783*. Chapel Hill: Univ. of North Carolina Press.
- DYER, MURRAY 1959 *The Weapon on the Wall: Rethinking Psychological Warfare*. Baltimore: Johns Hopkins Press.
- FARAGO, LADISLAS; and GITTNER, LEWIS F. (editors) 1942 *German Psychological Warfare*. New York: Putnam. → Contains an extensive list of references to German language material published before 1939.
- FULLER, J. F. C. 1920 *Tanks in the Great War, 1914-1918*. London: Murray.
- HOLT, ROBERT T.; and VAN DE VELDE, ROBERT W. 1960 *Strategic Psychological Operations and American Foreign Policy*. Univ. of Chicago Press.
- HUNTER, EDWARD 1956 *Brainwashing: The Story of Men Who Defied It*. New York: Farrar.
- LASSWELL, HAROLD D. (1927) 1938 *Propaganda Technique in the World War*. New York: Smith.
- LENER, DANIEL 1949 *Sykewar: Psychological Warfare Against Germany, D-Day to VE-Day*. New York: Stewart.
- LENER, DANIEL (editor) 1951 *Propaganda in War and Crises: Materials for American Policy*. New York: Stewart.
- LINEBARGER, PAUL M. A. (1948) 1954 *Psychological Warfare*. 2d ed. Washington: Combat Forces Press.
- MOCK, JAMES R.; and LARSON, CEDRIC 1939 *Words That Won the War: The Story of the Committee on Public Information, 1917-1919*. Princeton Univ. Press.
- Psychological Warfare and How to Wage It. 1940 *Current History and Forum* 51, no. 5: 52-53.
- SMITH, BRUCE L.; LASSWELL, HAROLD D.; and CASEY, RALPH D. 1948 *Propaganda, Communication and Public Opinion: A Comprehensive Reference Guide*. Princeton Univ. Press.
- THOMSON, CHARLES A. H. 1948 *Overseas Information Service of the United States Government*. Washington: Brookings Institution.
- YÜ, TÊ-CHI 1964 *Mass Persuasion in Communist China*. New York: Praeger.
- ZEMAN, Z. A. B. 1964 *Nazi Propaganda*. Oxford Univ. Press.

#### PSYCHOLOGY

Five major fields of psychology are described under this heading; other major fields are described under CLINICAL PSYCHOLOGY; COUNSELING PSYCHOLOGY; EDUCATIONAL PSYCHOLOGY; LEARNING; PERCEPTION; PERSONALITY; SOCIAL PSYCHOLOGY. The subject matter traditionally referred to as experimental psychology is found in the article on PHYSIOLOGICAL PSYCHOLOGY under this heading, as well as in DRIVES; EXPERIMENTAL DESIGN; FORGETTING; LEARNING; MOTIVATION; PERCEPTION; THINKING. Important and large-scale psychological theories are discussed in FIELD THEORY; GESTALT



THEORY; LEARNING, *articles on* CLASSICAL CONDITIONING and INSTRUMENTAL LEARNING; LEARNING THEORY; PSYCHOANALYSIS; THINKING, *article on* COGNITIVE ORGANIZATION AND PROCESSES.

I. COMPARATIVE PSYCHOLOGY	Roland H. Waters and Bradford N. Bunnell
II. PHYSIOLOGICAL PSYCHOLOGY	Clifford T. Morgan
III. CONSTITUTIONAL PSYCHOLOGY	W. L. L. Rees
IV. EXISTENTIAL PSYCHOLOGY	Rollo May and Sabert Basescu
V. APPLIED PSYCHOLOGY	Anne Anastasi

## I

## COMPARATIVE PSYCHOLOGY

Comparative psychology is concerned with the behavior of different species of living organisms. While some work is occasionally done with plant life, the field is typically limited to the behavior of animals. Its investigations lead to the specification of similarities and differences in behavior between species. These findings are of interest because it is assumed that they are correlated with the animals' positions on the phylogenetic scale. Interspecific comparisons should thus provide a means of relating an animal's behavior to its evolutionary background and contribute an important kind of information to psychology.

The accomplishments of the field have not, thus far, completely matched expectations. During the past century a vast amount of data on the behavior of infrahuman animals has been accumulated, but only a small fraction of this information has been used, or indeed proven useful, for comparative purposes. This has led some to deny the existence of comparative psychology as a distinct content area, and many more have used the title simply as the name for a method of study.

Since World War II, interest in comparative psychology has increased sharply. Some of the current trends in thinking and research that are beginning to revitalize comparative psychology will be presented in the sections below. For the purpose of this introduction, however, a few comments on some of the major characteristics of the field are in order.

(1) The term *comparative psychology* has often been used indiscriminately to include all work done on the behavior of infrahuman animals. Such animals are frequently used in behavioral studies because it is impractical, for economic or humanitarian reasons, to use human subjects and because the use of animals may be necessary in order to achieve an acceptable level of experimental control. There is nothing inherent in this practice

to justify calling it comparative psychology. Such studies typically have as their objective the analysis of a particular behavioral process, the animal subject being nothing more than a convenient vehicle for that behavior. The assumption implicit in this procedure is that the principles of behavior found in a particular species are relevant for all animals. This assumption is tempered somewhat by the circumstances that mammals are commonly used in such studies and that generalizations are usually made only to other mammals. Nevertheless, the assumption remains, even though it is this very assumption that comparative psychology is supposed to test. The point is, then, that some of what has been called comparative psychology is not only *not* comparative psychology but, in addition, assumes the validity of the evolutionary hypothesis on which the rationale for a comparative psychology is based.

(2) A second characteristic of comparative psychology is related to the first: from the beginning, comparative psychology has been highly anthropocentric in its orientation. In spite of the fact that man may be looked upon as just another species in the study of the evolution of behavior, he has not often been viewed in that way. Instead, comparative psychology has concentrated on the goal of achieving a better understanding of man through the study of the behavior of lower animals. This goal, which provided the original *raison d'être* for comparative psychology in Darwin's time, has continued to overshadow all others throughout the history of the field.

(3) The application of comparative methodology in comparative psychology has been very different from its application in other biological fields. For example, in classical zoological work concerned with the phylogenetic development of structure and function, comparisons are typically made between species within a genus. Except in the growing field of behavior genetics, little *behavioral* work has been done within genera and interspecific comparisons are usually made only across broad taxonomic groups—for example, a species of fish is compared with an amphibian, a reptile, or a mammal. The comparative psychologist's procedure is thus quite gross; he tends to select animals that are representative of relatively large morphological differences along the phylogenetic continuum. Therefore, the relationships sought by the comparative psychologist between an animal's evolutionary background and its behavior are of a very general nature. He justifies this approach by pointing out that his major interest is in the causal

analysis of behavior and only secondarily in the origins of behavior. Most of his data are relevant only to the former question. This approach is thus quite different from that of the ethologist, whose interest extends beyond causal analysis to include a primary concern with the evolutionary origins and adaptive significance of behavior. In ethology intrageneric comparisons have been used successfully in the identification and study of homologous behavior patterns.

(4) Finally, since the beginning of the twentieth century, comparative psychology has limited itself mainly to laboratory investigations. This practice developed out of a reaction against the anecdotal nature of much of the data collected during the latter half of the nineteenth century. These attempts to put behavior in a test tube, and to achieve rigorous experimental control, have been successful in producing reliable data and in reducing the teleological interpretations made by many early investigators. Comparative psychologists have been criticized on the grounds that many of their behavioral testing procedures, since they ignore the natural history of their subjects, do not provide an adequate sampling of the animal's behavioral repertoires and may actually yield spurious results because they impose conditions that strain or exceed the limits of the animal's capacities to respond. In response to this criticism, there has been a trend toward obtaining more and better information about an animal's behavior capacities before bringing it into the laboratory. The emphasis on laboratory experimentation continues, however, and the use of field studies and investigation of behavior under conditions and situations designed to give the animal full opportunity to express his "natural" responses serve to complement, rather than to supplant, the laboratory work.

### Historical aspects

It has been pointed out that man's interest in animal behavior is very nearly as old as man himself and that such information may well have been essential in order for man to eat and to avoid being eaten (Hale 1962). From a philosophical point of view, Aristotle taught that human and infrahuman behavior can be directly compared, that such characteristics as "gentleness or fierceness, mildness or cross temper, courage or timidity, fear or confidence, high spirit . . . or low cunning, and . . . sagacity" can be observed in animals below man (*Historia animalium* 588a). Such a belief was almost certainly a product of Aristotle's rational approach to man and nature rather than the re-

sult of the application of the observational techniques he defended in other connections. Aristotle's view yielded to the doctrine of "special creation" and the belief in qualitative differences between animals and men. It was not until this latter belief was challenged by the Darwinian theory of evolution that the seeds were sown for the development of contemporary comparative psychology.

The Darwinian theory denied the "special creation" explanation of the origin of man, erased the Cartesian dichotomy between besouled man and the soulless and mindless animal, and thus elevated the importance of animal study as a source of data for supporting evolutionary hypotheses. In order to demonstrate evolutionary continuity, it was necessary to show that man possessed many of the simpler, "brute" instincts present in lower animals and to show that these animals, in turn, exhibited capacities for reasoning, emotional behavior, social organization, and communication. That man and animals possess the simpler forms of behavior in common was never a matter of serious controversy, although later debates over the validity of the instinct concept are related to this question. Of greater importance was the search for higher mental processes in infrahuman animals. Examples of the first attempts to gather this sort of data are to be found in the writings of G. J. Romanes (1883). Such materials were largely anecdotal, however, and were harshly criticized because neither the validity of the observations reported nor the adequacy of the resulting interpretations could be checked by other workers. Demonstration of evolutionary principles was more a matter of anthropomorphic argument than scientifically derived fact.

Within experimental psychology, Wundt had championed the comparison of the behavior of man and animals as a means of better understanding man ([1864-1865] 1907, pp. 11, 340). Wundt thus threw the weight of his authority and position in psychology behind any future movement making the psychology of animals respectable. [See WUNDT.]

It was in this context that comparative psychology emerged as a highly anthropomorphic, anthropocentric area of investigation. The first major modification of this approach occurred in the 1890s. Loeb's tropism theory (1918), while not denying the animal mind, represented a return to the mechanistic view of Descartes. In 1894 C. Lloyd Morgan's "canon" appeared. This was a formulation of the principle of parsimony for the interpretation of animal behavior: "In no case may



we interpret an action as the outcome of the exercise of a higher psychical faculty, if it can be interpreted as the outcome of the exercise of one which stands lower in the psychological scale" ([1894] 1906, p. 53).

Together, Loeb and Morgan may be regarded as the fathers of scientific comparative psychology. Their writings stressed an objective approach and pointed the way to the use of experimental procedures in gathering data on animals (Warden 1927). It is important to note, however, that while anecdotalism was eliminated, the questions under study changed very little. Despite the new experimental fervor, investigators were still primarily concerned with describing the animal mind or consciousness, and with charting its evolutionary course. It was not until the arrival of behaviorism, some twenty years later, that the subject matter of comparative psychology shifted from the study of consciousness and volitional choice to the study of behavior *qua* behavior.

In the meantime, another development had occurred which was to have an important influence on the developing field. Thorndike, in his *Animal Intelligence: Experimental Studies*, came to the conclusion that all vertebrates "manifest fundamentally the same sort of intellectual life" ([1898-1901] 1911, p. 283). He found no important differences in the manner in which fish, reptiles, birds, and mammals learned the solutions to problems. The difference between a fish and a monkey was quantitative rather than qualitative—the monkey could form more complex associations more quickly and could retain them better than the fish. Man's intellect was said to be a further elaboration of the same sort of associative processes. This hypothesis was emphasized by Watson, and many came to accept the idea that the laws of learning were fixed and that there were no qualitative discontinuities in animal learning. This belief justified the use of rats and other infrahuman animals as human surrogates and tended to discourage comparative research. [See WATSON; THORNDIKE.]

Not everyone accepted Thorndike's conclusion, and some argued that although animals might not differ much in the capacity for simple associative learning, they might show significant differences on more complex tasks involving ideational or symbolic processes. Animals were tested on delayed response, double alternation, multiple choice, insight, and other such problems. Unfortunately, difficulties in providing standardized test situations and finding methods of quantifying the data were never resolved. By 1940 this approach had about died out. In consequence, at least in the United

States, the study of animal behavior no longer carried the flavor of a comparative approach.

A paper by Beach (1950) appeared as a protest against the then unsatisfactory status of comparative psychology. It had become almost exclusively limited to studies of the animal, with few, if any, applications of comparative methodologies. Animals were used as tools, or instruments, for the testing of different hypotheses about behavior that could not be tested on human subjects.

With the publication of Tinbergen's *The Study of Instinct* (1951) psychologists in the United States became increasingly aware of studies conducted by European scholars who referred to their area of research as ethology. The rapid postwar development of the field of ethology exerted a most important influence on the course of comparative psychology. These students of animal behavior were not interested in stopping with causal analyses of behavior. Their major objective was the determination of the way in which behavior evolved, and to this end they stressed the adaptive significance of certain species-specific behavior patterns. Presumably behavioral differences can arise in two ways: through genetic transmission during the evolution of the species and through learning during the ontogeny of the individual. For the ethologists, the most important behavioral elements were the highly stereotyped, relatively unmodifiable motor patterns, which were presumed to be innate and which were held to be the key to the evolution of behavior. These concepts, and the theories of instinctive behavior that were derived from them, clashed violently with the ideas of American psychologists who emphasized the importance of environment and experience. The past ten years, beginning with Lehrman's criticism (1953) of Lorenz' instinct theory, have witnessed a lively international debate between psychology and ethology. [For a typical rebuttal by the ethologists, see Eibl-Eibesfeldt & Kramer 1958, pp. 204-207; see also ETHOLOGY.]

Both fields have benefited from this debate. Following the lead of the psychologists, ethologists are now showing serious concern with the ontogeny of behavior, while ethology has stimulated psychology to a renewed interest in the field of behavior genetics. All the differences between the two have not yet been resolved. Arguments concerning the innateness of behavior, for example, may be expected to persist for some time.

These two events, Beach's protest on the negative side and the work of the ethologists on the positive side, have helped to revive interest in comparative psychology among American students.

### Evolution—structure and function

An increase in behavioral capacity presumably accompanies an increase in anatomical complexity. This generalization is usually true, although in some cases the presence of structural elements may have no apparent behavioral significance. For example, despite the presence of color receptors in the eye of the cat, it has been extremely difficult to demonstrate that this animal has functional color vision. Furthermore, an increase in behavioral complexity may appear without any apparent accompanying structural changes. Resolution of these problems will be, at least in part, dependent upon the development of better tests of behavioral "complexity" and upon improvement of anatomical, physiological, and chemical techniques applicable to the identification of structural changes in development.

With respect to the evolution of behavior, the problem of relating structure and function is difficult when function is defined in terms of the molar classifications that behavior psychologists generally use, such as learning, motivation, perception, etc. (In contrast, ethologists have emphasized a molecular behavioral unit, the "fixed action pattern," in their more successful attempts to trace the evolution of some aspects of behavior: e.g., Lorenz 1958.)

Since behavior does not exist in fossil form, an additional difficulty is that evidence for the evolutionary course of behavior is necessarily indirect. It must also be recognized that the behaviors available for study are exhibited by highly specialized creatures that structurally, and presumably behaviorally as well, differ considerably from the ancestral source of the evolutionary branch. [See *EVOLUTION*, especially the article on *EVOLUTION AND BEHAVIOR*.]

Comparative psychologists have tended to emphasize central, as opposed to peripheral, mechanisms. Within any broad taxonomic group, such as a phylum, or even a class, there typically exists a tremendous range of sensory and motor development. For example, the visual systems of coelenterates range from pigmented dermal spots to relatively sophisticated lens eyes (Maier & Schneirla 1935). Across phyla there are many instances in which a particular representative of a "lower" position may be better equipped for receiving a particular kind of energy than a representative of a "higher" group; for example, many arthropods have color vision, while many mammals apparently do not. While a number of investigators have been interested in the functional correlates of different

types of peripheral systems, those psychologists who study learning, motivation, and perception stress the role of the central nervous system and typically concern themselves only with making certain that sensory information is reaching the animal and that the overt responses being measured are within its behavioral repertoire. Their interest is in how the animal uses the information it receives and in what effect the information has on the response system. They usually assume some sort of intervening process, or processes, mediated by the internuncial structures of the organism. "Intelligence" is typically thought of as such a central process that is highly correlated with the level of neural organization possessed by the organism. This is the working hypothesis behind much of American comparative psychology [see "Learning," *below*; see also *NERVOUS SYSTEM*, article on *STRUCTURE AND FUNCTION OF THE BRAIN; SENSES*.]

These factors go a long way toward explaining the fact that in comparative psychology comparisons are typically made across broad groups, a procedure that is not employed by the zoological taxonomist and one that is not very useful for the delineation of the phyletic history of a given bit of behavior. The procedure may be justified, however, on the grounds that these are comparisons between representatives of different levels of neural organization and that this is a most useful type of information, particularly when one considers the categories of behavior in which the psychologist is most typically interested.

### Instinct, ontogeny, and motivation

**Instinct.** Beginning in the 1920s the entire discipline of psychology underwent what has been called the "anti-instinct" revolt (Beach 1955). The major objection was to the use of instinct as an explanatory concept. It was pointed out that calling a behavior an instinct did not in any way demonstrate the causes of that behavior and that the use of such a label tended to keep people from doing the research necessary to determine causal factors. It was felt that if such research were done, the behavior would prove amenable to analysis in behavioristic terms and that the instinct concept was superfluous (Kuo 1924). In support of this position, it was possible to point to work such as that of Kuo (1930), who analyzed the rat-killing "instinct" of the cat and showed it to be heavily dependent upon the past history of the organism. The cat's behavior, with respect to rats, could be modified extensively, and a wide range of responses to rodents could be developed. This and similar studies fit well with the contemporary behavioristic



emphasis upon behavioral plasticity and the importance of experience and learning. In terms of the interests of psychologists, many "instinctive" behaviors could be considered as little more than special applications of the principles being generated in the learning laboratories. The net result of all this was to banish the instinct concept from psychology. Along with it, with a few significant exceptions, went most of the interest in behavior genetics and in the study of behavioral constancy. Conceptually, the essential importance of accepting an epigenetic interpretation of instinctive behaviors, an interpretation that stresses the *interaction* between heredity and environment, was recognized (Carmichael 1927). In actual fact, psychologists emphasized experiential factors to the exclusion of almost everything else.

In the 1950s the debate over the instinct concept arose once more, stimulated in large part by the work of the ethologists. The central question again was whether or not *instinct*, if there was such a thing, was a meaningful class of behavior. Instinctive behavior, defined essentially as the product of innate neural patterns laid down by the genetic constitution of the organism, was the basic behavioral datum for the program of the ethologists. Given the evolutionary interests of ethology, behavior that was typical of a species commanded its greatest attention. [See INSTINCT.]

American psychologists once again reacted negatively to the instinct concept and again defended the hypothesis of an interaction between genetic constitution and experience; "instinct" could not be either learned or inherited. As has been noted, the net result of the entire debate was a shift of emphasis to studies of the ontogeny of behavior and a reawakened interest in behavior genetics. [See GENETICS, *article on* GENETICS AND BEHAVIOR.]

An example of the contemporary American approach to the study of instinctive, or species-typical, behavior patterns is the work of Lehrman on the analysis of the reproductive behaviors of the ring dove (for summaries of this work, see Lehrman 1932; 1964). He has been able to demonstrate experimentally several of the mechanisms by which hormonal-environmental interactions give rise to the expression of complete behavior patterns. In doing this, Lehrman (1962) has called attention to the fact that experience may have an effect on behavior—by altering rates of growth, producing changes in sensitivity, facilitating behaviors that occur later in a sequence, etc.—which does not fit the standard definitions of "learning," but which nevertheless can be crucial for the normal develop-

ment of the behavior pattern. The generalization is drawn that learning and heredity cannot be meaningfully distinguished, and the traditional view that learning is dominant in higher animals and instinct in lower animals is no longer considered valid by most investigators. Some, however, still find the dichotomy useful (e.g., Thorpe 1956). Schneirla puts the situation in these words:

... heredity influences all behaviour in all animals, its influence is more direct in some animals (evidently the more primitive) than in others and may be very different on different phyletic levels. . . . [It is] likely that behaviour evolves through new and increasingly complex reorientations, with the old rebuilt and extended in terms of the new, [rather] than through the replacement of the "innate" by the "learned."

... [It] is therefore probable that the relationship of behaviour to survival, always important in evolution, is very different on various phyletic levels. As it is realized that assumptions of innateness in opposition to the acquired can explain no activity, increasing emphasis is placed upon the need for analytical investigation. (1963, p. 692)

**Ontogeny.** In addition to the ontogenetic approach, which constitutes the mainstream of American comparative psychologists' work with species-typical behavior patterns, there is a continuing interest in ontogenetic development for its own sake. This is reminiscent of the traditional stress by Freud and Watson on the effects of the early history of the organism.

Beach and Jaynes (1954) cite the following examples of animal studies in this area that serve as an indication of the scope of the work on the effect of early experience: rats, pigeons, and chimpanzees, raised in darkness, exhibit marked but not permanent disturbances in later visual adjustments; the appearance of normal motor activities in chimpanzees raised under conditions of restricted movement is temporarily delayed; chickens raised in the dark and prevented from pecking are less accurate in pecking but rapidly improve with practice; some evidence of a "need to suck" is reported when puppies are deprived of the sucking activity; the effects that depriving infant rats of food and water has on their hoarding of food and water as adults are somewhat ambiguous; egg-laying behavior in some species of insects can be elicited by other than the normal host when they are raised on a nonpreferred host; mating behavior in certain species of fish, ring doves, and pigeons can be elicited by a species different from their own when the former are raised alone or with the foreign species; reproductive behavior of such mammals as rats, the American bison, antelopes, and

others is modified under conditions of captivity; species recognition and gregariousness in insects, birds, and domestic chicks are subject to modification through early experiences, as is song acquisition in birds; social behaviors in the dog, in the lambs of both wild and tame sheep, and in chimpanzees are all modifiable during early infancy; wildness and aggressiveness, and the ability to adapt to differing learning situations, in rats have been modified by "gentling" procedures and by varying the type of environmental conditions present during infancy.

Contemporary examples of research in this area, to mention only a few, are the work of Scott and Marston (1950) on the effects of early experience on the development of adult behavior in dogs, which called attention to the presence of critical periods in that development; the studies of imprinting, which also emphasized the critical-period concept (see Hess 1959); the studies of stress effects (e.g., Levine 1962); and Harlow's (1958) work on affectional responses in infant monkeys.

A model proposed by Fuller and Waller (1962) brings together some of the explanations that have been proposed to account for early experience effects. One set of explanations rests on the concept of learning: ability to learn a specific type of response may change with age, a change that cannot at present be correlated with any known neurophysiological change; it may also be that what the results show is the persistence of early habits resulting from such conditions as change in drive, amount and kind of reinforcement, lack of interfering or competing habits, or similar influences.

A second set of explanations rests on the concept of maturational changes. Such changes may be inferentially applied to the learning changes given above; they may contribute to the structural and neurophysiological conditions underlying sensory, neural, and muscular equipment employed in specific responses because of stimulation during early experience; and, lastly, the finding of so-called critical periods in development, times during which mastery of particular acts or the effect of stimulation is more marked, may be a reflection of the fact that the systems supporting such activities have reached maturity.

A third set of concepts is couched in terms of the relatively greater sensitivity of the young organism to all sorts of physical and chemical variation. Presumably, then, they are more susceptible to nonspecific stresses in early life and might as a result undergo permanent modifications in basic physiological functions. Such alterations in meta-

bolic, autonomic, and endocrine systems would be expected to produce variations in adult behavior patterns. [See INFANCY.]

**Motivation.** The approach to many of the problems of the energizing and directive processes operative in such maintenance activities as feeding, drinking, and reproduction has already been mentioned. The current emphasis is upon the analysis of the entire behavior pattern, rather than on the measurement of the amount of consummatory activity alone. Aronson (1959), for example, has pointed out that phylogenetic trends in the relationships between sex hormones and reproductive behaviors are often discernible only when careful attention is paid to the individual components of the behavior pattern. When motivation is considered in terms of consummatory behavior, as Ratner and Denny (1964, chapter 9) point out, the higher an animal is on the phylogenetic scale, the greater is the variety of stimulation that can elicit or control the behavior. Studies of the sexual behavior of male mammals below the primates (e.g., Hale & Almqvist 1956; Fisher 1962) have shown that ungulates, and even rodents, are very responsive to stimulus variation. A similar trend with regard to the operation of experiential effects is illustrated by work on cats (Rosenblatt & Aronson 1958), rats (Zimbardo 1958), and guinea pigs (Valenstein, Riss, & Young 1955). While this work generally supports the notion that increasing behavioral flexibility accompanies increases in phylogeny, perhaps the most interesting finding of all has been that environmental and experiential influences are more ubiquitous and are of critical importance further down the scale than had previously been thought.

In the past, feeding, drinking, and sleeping have been investigated more by zoologists and physiologists than by comparative psychologists. Newer methods of approach, particularly those which have made the central nervous system accessible to experimental approach, now seem to be leading to increased interest in comparative studies of these fundamental processes.

Beyond the study of basic maintenance functions, the work of Butler (1953) on exploratory drives and of Harlow and his associates (e.g., Harlow, Harlow, & Meyer 1950; Harlow & McClearn 1954) on the manipulative drive in monkeys suggest evidence for "supraphysiological" drives, at least in higher animal forms. Berlyne (1960) provides a good discussion of curiosity and investigative motives as "higher" drives. Finally, the idea that some drives may be learned (Miller



1951) and that motivated behavior may be controlled by secondary reinforcement mechanisms is widely accepted. Presumably such learning is more characteristic of the higher animals. [See DRIVES; MOTIVATION; STIMULATION DRIVES.]

### Learning

The field of learning has traditionally been the major area of interest in American comparative psychology. In fact, this emphasis has resulted in the strongest criticism of comparative psychology by other students of animal behavior. Many feel that significant aspects of the animal's behavior are often missed completely or are given an erroneous interpretation because of the concentration on only one facet of behavior.

There are several problems involved in the learning area. The first is the definition of learning itself. The emphasis in much of the work of comparative psychologists has been upon classical conditioning and instrumental learning paradigms. The establishment of a link between a stimulus and a response is, however, only one of the ways by which organisms modify their behavior as a result of environmental change or as a consequence of their past activities. Many of the techniques by which the organism adapts to its environment are not "true" learning. Fatigue effects represent one such example. While fatigue may not be hard to understand, the phenomena of sensitization and habituation, which typically have been controlled out of laboratory learning situations as being unwanted "noise," cannot be disregarded so easily. These types of behavioral modification appear to be universal in the animal kingdom and deserve more attention than they have been given as possible precursors of true associative learning. It is probable that as psychologists conduct more studies on the invertebrates, these phenomena will be more carefully analyzed. Such emphasis is deserved since, at the mammalian level, studies (reviewed in Livingstone 1959) have clearly demonstrated central nervous control of peripheral sensory processes, which appears to be a critical mechanism in habituation. Other special phenomena, such as provisional orientation (Maier & Schneirla 1935), imprinting (Hess 1959), and the type of behavioral modification shown by Lehrman's ring doves, are being given the increased attention they merit. [See FATIGUE; IMPRINTING; LEARNING.]

A second problem is the question of the lowest phyletic level at which learning occurs. This problem has been attacked using the traditional paradigms for associative learning. At present it is certain that arthropods and at least some mollusks

are capable of such learning. Also, the evidence is good that annelids and flatworms can be trained by means of both classical conditioning and instrumental techniques (Jacobson 1963). Below this level reports of learning are not convincing, although the reader is referred to the work of Gelber (1952) and of Jensen (1957) for a debate over the presence of associative learning in one-celled animals. Questions for future study will be: whether there is any difference in the relative primitiveness of classically, as opposed to instrumentally, conditioned responses; and whether the various types of learning, as now defined, are mediated by the same neural and/or chemical processes. It is at least possible that the same result, in terms of a learning proficiency score, is obtainable through more than one mechanism.

A third problem is the search for the relationship between phyletic position and learning capacity. In general, it has been found that as one ascends the phylogenetic scale there is a corresponding increase in the ability of organisms to modify their behaviors when they are subjected to environmental changes. This behavioral plasticity is revealed by the fact that "higher" animals are able to learn more complex problems than are "lower" forms. However, this generalization has frequently broken down, for there are many reports in which no apparent differences have been revealed between representative phyla or classes, and some instances in which "lower" forms surpassed "higher" forms in performance on learning tasks. It is this sort of finding that challenges the evolutionary hypothesis of behavioral development. It does not disprove such a hypothesis, but it does mean that close attention must be paid to the situations in which the tests are conducted and to the methods used in obtaining the scores that are to be compared. [See LEARNING, *article on LEARNING IN CHILDREN.*]

A partial resolution of conflicts between evolutionary theory and data is found when one explores the proposition that certain "lower" species may have evolved highly specialized abilities to learn particular things in particular situations. It is part of the task of the comparative psychologist to recognize such occurrences and to try to identify the mechanisms that mediate their appearance.

The acknowledgment of species-specific learning capacities does not help very much in demonstrating a general relationship between phyletic position and learning capacity, however. As has been noted, Thorndike's failure to find differences in the problem-solving performance of many different species diverted many from the use of comparative methods. The finding that if an animal could do a sim-

ple learning task at all, it could do it about as well as any other animal was repeatedly corroborated.

Some investigators, however, felt that differences between species might become apparent if more complex tests of animal "intelligence" were used. Delayed-response, multiple-choice, and double-alternation problems were among the many early tests devised to tap the higher mental processes of animals. More recently, such techniques as complex multiple-cue tasks (such as oddity, matching, and conditional-discrimination problems), probability-matching techniques, and both reversal learning-set and discrimination learning-set problems have been introduced.

In the interpretation of results of comparative studies of learning a useful distinction can be made between quantitative and qualitative comparisons of performance. Almost all the early work consisted of attempts at quantitative comparison. Typically, scoring was on a continuum of some sort (number of seconds of delay possible, number of errors, running time, trials to criterion, etc.), and the idea was that different species ought to fall at different places on such a continuum. Since interspecific differences in sensory and motor capacities, to say nothing of motivational-emotional factors, are frequently considerable, the investigator who uses such an approach is faced with the difficult task of defining standard test situations and scoring procedures.

A study by Schneirla (1946), in which the maze performance of rats and ants was compared, provides an illustration of what is meant by qualitative comparison. Schneirla demonstrated that rats learned more quickly than ants (a quantitative difference), but he emphasized his discovery that the way in which ants went about learning the maze was very different from the processes exhibited by the rats. Ants learned one section of the maze at a time, and the learning of one section resulted in an increase in errors on succeeding sections, as if the maze were "new" after each portion of the maze was completed. In addition, the ants eliminated blind alleys in the maze differently than did the rats, making successively shallower penetrations into cul-de-sacs while the rats tended to eliminate each wrong entry in an all-or-none fashion. When the maze was reversed after learning had been completed, the rats showed considerable savings, but ants acted as though the situation were totally new and had to be relearned from the beginning. Finally, ants did not show any goal gradient effect (earlier elimination of the blinds near the goal), while the rats did. Thus, a meaningful differentiation between the two species could

be made on qualitative grounds in terms of how the animals went about learning and of what was learned. Bitterman (1960) and his associates have been among the first to adopt this qualitative approach. They have, for example, shown that the "Humphreys effect" (greater resistance to extinction after training under partial, as compared with continuous, reinforcement) does not occur in fish in the same manner as it does in rats. This indicates that, for fish, the relationship between the outcome of a response (its reinforcement) and the response is much more direct, and presumably simpler, than in higher vertebrates.

A current trend in the comparative psychology of learning thus emphasizes a search for qualitative differences—a comparison of animals in terms of the extent to which they follow the same principles in their learning and whether the functions plotted from raw scores exhibit similar shapes or other characteristics. Warren (1965) has provided a comprehensive summary of both qualitative and quantitative comparative data on vertebrate learning performance.

A final problem, cutting across all of the problems mentioned so far, is that of finding ways of equating testing situations so that comparisons can be made. As was noted earlier, the experimenter cannot be sure that in a "standard" situation an animal of one species can receive the same sensory information as one of another species, can perform the response with equal facility, or is similarly motivated. Failure to control these conditions can prevent any meaningful comparisons. Unfortunately, except perhaps with animals that have very similar sensory and motor capacities, it is difficult to control these factors by equating the tasks at hand, and even were this done, the problem of equating motivational state and controlling for different ecological backgrounds is extremely difficult. The way around these difficulties, as proposed by Bitterman (1960), is to employ the method of control by variation. If, for example, the absence of the Humphreys effect in fish is duplicated under different motivational states, and over other testing situations, etc., it would be fairly safe to say that a meaningful difference exists between these animals—fish and rats—relative to this effect. It would not be necessary to argue that the animals were tested under "culture free" conditions.

### Social behavior

In its broadest sense "social behavior" refers to behavior elicited by and directed toward another individual or group. Such behavior is clearly interactive when the behavior of the first individual is



modified by the response of the second individual or group. Social behavior as thus defined may be readily observed taking place between individuals of different species: the shepherd's dog responds to signals given by his master as well as to the movements of the members of the flock; a predatory animal elicits "alarm" and "escape" or "defense" reactions from its potential prey; the mockingbird drives birds of another species away from its territory; and so on, with numerous other instances of interaction between species and between genera. More critical and significant for the purpose of the present review, however, is the social behavior taking place within species. It is from these studies that principles of group organization and types of social relationships are being uncovered.

A comparative psychology of social behavior should be interested in such general questions as: Where, on the phylogenetic scale, does social behavior first occur? What is the relationship between position on the phyletic scale and the types of social behavior observed? On what bases may a classification of social behavior be made so that such comparisons are possible for both phylogenetic and ontogenetic development?

With regard to the first question, one attempt at a solution has been to focus on the concepts of cooperation and competition. The animal sociologist W. C. Allee (1931; 1938) believed that even among simple organisms there exists an automatic, unconscious "protocooperation," which is a fundamental biological principle. This principle says that at the level of the social group the interrelations between individuals in the group are more helpful than harmful. True social behavior is thought to have developed from animal aggregations which, while initially formed by nonsocial forces, survived because of advantages obtained through proximity. According to Allee, subsequent development followed a familiar pattern: "Each advance in complexity came from the natural selection of an increased ability in natural cooperation on the part of the evolving stock; the greater natural cooperation came first, then it was selected. Cooperation is a biological necessity both in individual evolution and in the evolution of groups" ([1938] 1951, p. 211). One very important factor was the evolution of the cooperative functions involved in sexual reproduction. Allee (1931) thought it likely that sex originally arose from interorganismic stimulation produced as a consequence of the aggregation of simple asexual organisms. The appearance of sexual reproduction thus provided for the appear-

ance of the family, one of the basic units in true social groupings and social behavior.

Closely allied with the concept of cooperation has been the problem of competition between individuals and the ways that have been developed for handling aggressive behavior (e.g., territoriality, dominance-subordination relationships, and the like). Aggressive behavior occurs in some of the higher invertebrates, but with few exceptions studies of aggression and fighting have been confined to vertebrates. The role of aggressive behavior in vertebrates has been described in a classic paper by Collias (1944). A more recent summary of concepts and data in the area of competition and cooperation (as well as papers on reproductive behaviors, socialization, and the evolution of signaling devices) can be found in the book *Social Behavior and Organization Among Vertebrates*, edited by Etkin (1964).

It is quite possible that principles other than those of cooperation and competition underlie the evolutionary development of social behaviors, but alternative conceptual and theoretical models are not available.

The question of the relation between phyletic level and the types of social behavior present has received little systematic study. There are a number of studies dealing with the social behavior of a given species, but they have not been ordered in a fashion that permits an approach to the determination of the relation between phyletic level and types of behavior exhibited. Not all investigators in the area have shown a desire for a classificatory system of social behavior, many apparently preferring to take the behavior as they see it without attempting to force it into rigid categories. It is clear, however, that some sort of classification is necessary if the search for phylogenetic relationships is to be taken seriously.

The schema proposed by Scott is an example of an attempt at such a classification, and an example of its application for comparative purposes is available (1960, pp. 261-264). Tinbergen (1953) has provided an analysis of social behavior from the ethological point of view.

Neither the Scott nor the Tinbergen classification in itself implies anything about the mechanisms by which social behavior is established. In actual fact, Tinbergen emphasizes the innateness of such behaviors, while Scott has stressed a "socialization" concept in which the effects of experience are presumed to play an important role. Insofar as flexibility and the ability to learn are related to the complexity of social responses, it may be suggested

that "higher" animals are capable of assuming a wider variety of roles and relationships and that these are more interchangeable than they are in "lower" forms. This generalization has not, however, been extensively tested. Finally, the development of language provides for a tremendous increase in the communications that are so essential to social behavior, giving a significant advantage in the potential for social behavior. [See SOCIAL BEHAVIOR, ANIMAL.]

ROLLAND H. WATERS AND  
BRADFORD N. BUNNELL

[See also ETHOLOGY. Other relevant material may be found in COMMUNICATION, ANIMAL; EVOLUTION; GENETICS; IMPRINTING; INSTINCT; SEXUAL BEHAVIOR, article on ANIMAL SEXUAL BEHAVIOR; SOCIAL BEHAVIOR, ANIMAL; and in the biographies of DARWIN and MORGAN, C. LLOYD.]

#### BIBLIOGRAPHY

- ALLEE, W. C. 1931 *Animal Aggregations: A Study in General Sociology*. Univ. of Chicago Press.
- ALLEE, W. C. (1938) 1951 *Cooperation Among Animals, With Human Implications*. Rev. & enl. ed. New York: Schuman. → First published as *The Social Life of Animals*.
- ARONSON, LESTER R. 1959 Hormones and Reproductive Behavior: Some Phylogenetic Considerations. Pages 98-120 in Columbia University Symposium on Comparative Endocrinology, Cold Spring Harbor, N.Y., 1958, *Comparative Endocrinology: Proceedings*. New York: Wiley.
- BEACH, FRANK A. 1950 The Snark Was a Boojum. *American Psychologist* 5:115-124.
- BEACH, FRANK A. 1955 The Descent of Instinct. *Psychological Review* 62:401-410.
- BEACH, FRANK A.; and JAYNES, JULIAN 1954 Effects of Early Experience Upon the Behavior of Animals. *Psychological Bulletin* 51:239-263.
- BERLYNE, D. E. 1960 *Conflict, Arousal, and Curiosity*. New York: McGraw-Hill.
- BITTERMAN, M. E. 1960 Toward a Comparative Psychology of Learning. *American Psychologist* 15:704-712.
- BUTLER, ROBERT A. 1953 Discrimination Learning by Rhesus Monkeys to Visual-exploration Motivation. *Journal of Comparative and Physiological Psychology* 46:95-98.
- CARMICHAEL, LEONARD 1927 A Further Study of the Development of Behavior in Vertebrates Experimentally Removed from the Influence of External Stimulation. *Psychological Review* 34:34-47.
- COLLIAS, N. E. 1944 Aggressive Behavior Among Vertebrate Animals. *Physiological Zoology* 17:83-123.
- EIBL-EBESFELDT, IRENÄUS; and KRAMER, SOL 1958 Ethology: The Comparative Study of Animal Behavior. *Quarterly Review of Biology* 33:181-211.
- ETKIN, WILLIAM (editor) 1964 *Social Behavior and Organization Among Vertebrates*. Univ. of Chicago Press.
- FISHER, ALAN E. 1962 Effects of Stimulus Variation on Sexual Satiation in the Male Rat. *Journal of Comparative and Physiological Psychology* 55:614-620.
- FULLER, JOHN L.; and WALLER, MAREUS B. 1962 Is Early Experience Different? Pages 235-245 in E. L. Bliss (editor), *Roots of Behavior: Genetics, Instinct, and Socialization in Animal Behavior*. New York: Harper.
- GELBER, BEATRICE 1952 Investigations of the Behavior of *Paramecium aurelia*: 1. Modification of Behavior After Training With Reinforcement. *Journal of Comparative and Physiological Psychology* 45:58-65.
- HALE, E. B. 1962 Domestication and the Evolution of Behaviour. Pages 21-53 in E. S. E. Hafez (editor), *The Behaviour of Domestic Animals*. Baltimore: Williams & Wilkins.
- HALE, E. B.; and ALMQUIST, J. O. 1956 Effect of Changes in the Stimulus Field on Responsiveness of Bulls to a Constant Stimulus Animal. *Anatomical Record* 125: 607 only.
- HARLOW, HARRY F. 1958 The Nature of Love. *American Psychologist* 13:673-685.
- HARLOW, HARRY F.; HARLOW, M. K.; and MEYER, D. R. 1950 Learning Motivated by a Manipulation Drive. *Journal of Experimental Psychology* 40:228-234.
- HARLOW, HARRY F.; and MCCLEARN, GERALD E. 1954 Object Discrimination Learned by Monkeys on the Basis of Manipulation Motives. *Journal of Comparative and Physiological Psychology* 47:73-76.
- HESS, ECKHARD H. 1959 Imprinting. *Science* 130:133-141.
- JACOBSON, ALLAN L. 1963 Learning in Flatworms and Annelids. *Psychological Bulletin* 60:74-94.
- JENSEN, DONALD D. 1957 Experiments on "Learning" in *Paramecia*. *Science* 125:191-192.
- KUO, ZING YANG 1924 A Psychology Without Heredity. *Psychological Review* 31:427-448.
- KUO, ZING YANG 1930 The Genesis of the Cat's Responses to the Rat. *Journal of Comparative Psychology* 11:1-35.
- LEHRMAN, DANIEL S. (1932) 1961 Hormonal Regulation of Parental Behavior in Birds and Inframammals. Volume 2, pages 1268-1382 in W. C. Young (editor), *Sex and Internal Secretions*, 3d ed. Baltimore: Williams & Wilkins.
- LEHRMAN, DANIEL S. 1953 A Critique of Konrad Lorenz's Theory of Instinctive Behavior. *Quarterly Review of Biology* 28:337-363.
- LEHRMAN, DANIEL S. 1962 Interaction of Hormonal and Experiential Influences on Development of Behavior. Pages 142-156 in E. L. Bliss (editor), *Roots of Behavior: Genetics, Instinct, and Socialization in Animal Behavior*. New York: Harper.
- LEHRMAN, DANIEL S. 1964 Control of Behavior Cycles in Reproduction. Pages 143-166 in William Etkin (editor), *Social Behavior and Organization Among Vertebrates*. Univ. of Chicago Press.
- LEVINE, SEYMOUR 1962 Psychophysiological Effects of Infantile Stimulation. Pages 246-253 in E. L. Bliss (editor), *Roots of Behavior: Genetics, Instinct, and Socialization in Animal Behavior*. New York: Harper.
- LIVINGSTONE, ROBERT B. 1959 Central Control of Receptors and Sensory Transmission Systems. Volume 1, pages 741-760 in *Handbook of Physiology*. Section 1: Neurophysiology. Washington: American Physiological Society.
- LOEB, JACQUES 1918 *Forced Movements, Tropisms, and Animal Conduct*. Philadelphia: Lippincott.
- LORENZ, KONRAD Z. 1958 The Evolution of Behavior. *Scientific American* 199, no. 6:67-78.



- MAIER, NORMAN R. F.; and SCHNEIRLA, T. C. 1935 *Principles of Animal Psychology*. New York: McGraw-Hill.
- MILLER, NEAL E. 1951 Learnable Drives and Rewards. Pages 435-472 in S. S. Stevens (editor), *Handbook of Experimental Psychology*. New York: Wiley.
- MORGAN, C. LLOYD (1894) 1906 *An Introduction to Comparative Psychology*. 2d ed. New York: Scribner; London: Scott.
- RATNER, STANLEY C.; and DENNY, M. RAY 1964 *Comparative Psychology: Research in Animal Behavior*. Homewood, Ill.: Dorsey.
- RIOPPELLE, ARTHUR J. 1960 Complex Processes. Pages 208-249 in R. H. Waters, D. A. Rethlingshafer, and W. E. Caldwell (editors), *Principles of Comparative Psychology*. New York: McGraw-Hill.
- ROMANES, GEORGE J. (1883) 1910 *Animal Intelligence*. 9th ed. London: Routledge.
- ROSENBLATT, JAY S.; and ARONSON, LESTER R. 1958 The Decline of Sexual Behavior in Male Cats After Castration With Special Reference to the Role of Prior Sexual Experience. *Behavior* 12:285-338.
- SCHNEIRLA, T. C. 1946 Ant Learning as a Problem in Comparative Psychology. Pages 276-305 in P. L. Harriman (editor), *Twentieth Century Psychology: Recent Developments in Psychology*. New York: Philosophical Library.
- SCHNEIRLA, T. C. 1963 Comparative Psychology. Volume 18, pages 690-703 in *Encyclopaedia Britannica*. Chicago: Benton.
- SCOTT, J. P. 1960 Comparative Social Psychology. Pages 250-288 in R. H. Waters, D. A. Rethlingshafer, and W. E. Caldwell (editors), *Principles of Comparative Psychology*. New York: McGraw-Hill.
- SCOTT, J. P.; and MARSTON, MARY V. 1950 Critical Periods Affecting the Development of Normal and Maladjustive Social Behavior of Puppies. *Journal of Genetic Psychology* 77:25-60.
- THORNDIKE, EDWARD L. (1898-1901) 1911 *Animal Intelligence: Experimental Studies*. New York: Macmillan.
- THORPE, WILLIAM H. (1956) 1963 *Learning and Instinct in Animals*. 2d ed., rev. & enl. Cambridge, Mass.: Harvard Univ. Press.
- TINBERGEN, NIKOLAAS 1951 *The Study of Instinct*. Oxford: Clarendon.
- TINBERGEN, NIKOLAAS 1953 *Social Behaviour in Animals, With Special Reference to Vertebrates*. London: Methuen.
- VALENSTEIN, ELLIOT S.; RISS, WALTER; and YOUNG, WILLIAM C. 1955 Experiential and Genetic Factors in the Organization of Sexual Behavior in Male Guinea Pigs. *Journal of Comparative and Physiological Psychology* 48:307-403.
- WARDEN, C. J. 1927 The Historical Development of Comparative Psychology. 4: From the Sixteenth Century to Darwin. *Psychological Review* 34:135-168.
- WARREN, J. M. 1965 Primate Learning in Comparative Perspective. Pages 249-281 in Allan M. Schrier, Harry F. Harlow, and Fred Stollnitz (editors), *Behavior of Nonhuman Primates: Modern Research Trends*. New York: Academic Press.
- WUNDT, WILHELM (1864-1865) 1907 *Lectures on Human and Animal Psychology*. New York: Macmillan. → First published in German.
- ZIMBARDO, PHILIP G. 1958 The Effects of Early Avoidance Training and Rearing Conditions Upon the Sexual Behavior of the Male Rat. *Journal of Comparative and Physiological Psychology* 51:764-769.

## II

## PHYSIOLOGICAL PSYCHOLOGY

Physiological psychology is the study of the physiological basis of behavior and mental life. In other words, it is concerned with the role of various organs of the body in psychological phenomena. Of these organs, the brain is clearly the most important, for it is the great coordinator and organizer of mental and behavioral events. The subdivision of physiological psychology concerned exclusively with the brain is known as *neuropsychology*—the title of the chair held by Karl Lashley at Harvard from 1937 until 1955. But other organs of the body also play important roles. These fall into two main categories: the sensory (receptor) systems and the motor (effector) systems. The motor systems include not only the skeletal and smooth muscles but also the endocrine glands. These latter, through their various effects on sensory and motor organs as well as on the nervous system itself, are proving to play an important role in the determination of behavior. Their role is, of course, biochemical. Also biochemical and included in physiological psychology are the influences of diet and of drugs. The general field of chemical influences on behavior has come to be called *psychochemistry*. That which is restricted to the effects of drugs is *psychopharmacology*. [See DRUGS, article on PSYCHOPHARMACOLOGY; MENTAL DISORDERS, article on BIOLOGICAL ASPECTS; NERVOUS SYSTEM; SENSES.]

In principle physiological psychology covers both human and animal activities, but in practice it has been more concerned with animal behavior. This is because the techniques for studying physiological correlates in man are limited largely to clinical observations or measurements, and these are not so illuminating or powerful as the experimental methods that can be used with animals. For example, experimental surgery and both electrical stimulation and recording deep within the organs of the body are seldom possible with human beings, but they are with animals. Similarly, there are restrictions on the use of drugs and biochemical agents which apply to human beings but not to animals. The situation is changing somewhat as new techniques become available, and the special field that emphasizes physiological correlates of human psychological phenomena has recently been called *psychophysiology*.

**Relationship to other fields.** It should be noted that physiological psychology is related to three other fields: experimental psychology, psychobiology, and comparative psychology. Historically the term *experimental psychology*, introduced about

1880, meant the application of the experimental method to problems of psychology formerly addressed through philosophical reflection and observation. Today the term covers not only the method but the body of facts and principles acquired with the method. Since most of what is known in physiological psychology was learned from the experimental method, physiological psychology is largely a subdivision of experimental psychology.

*Psychobiology* was a term widely used in the first part of the twentieth century, and then it fell into disuse. It has recently been revived, particularly among biologists, with the flowering of interest in differences in behavior among species. Etymologically, psychobiology could very well apply to any and all relations between psychological and biological phenomena, including physiological psychology. But scientific usage, being determined by scientists rather than etymologists, has narrowed the term primarily to species differences in behavior. Thus, psychobiology refers to the study of the behavior often observed in natural settings or in quasi-experimental situations as it relates to the biological and genetic characteristics of the species under study rather than to the functions of internal organs of the body. The term *ethology* has been used by European zoologists to refer to this kind of psychobiology.

In the United States the term *comparative psychology* has come, at least among psychologists, to displace psychobiology. Comparative psychology, as the term implies, is supposed to be a comparison of the behavior of different species along the phyletic scale. However, because of the convenience and cheapness of the laboratory rat, so much of American comparative psychological research, until recently, was done with the rat that the term "comparative" was hardly justified. The situation has changed somewhat, but American experimental work with animal behavior is still better characterized as *animal psychology* than as comparative psychology. The ethology of the European zoologists is closer to the meaning of comparative psychology, even though it is more naturalistic and less exact than experimental animal psychology. [See ETHOLOGY and PSYCHOLOGY, article on COMPARATIVE PSYCHOLOGY.]

**Methods.** Physiological psychology employs the methods of several fields but principally, of course, those of psychology and physiology. A typical experiment or study consists of correlating a psychological variable and a physiological variable. That is to say, any psychological experiment of interest may be combined with some physiological inter-

vention. Since psychological experiments are well covered in other articles in this encyclopedia, only physiological methods will be discussed briefly here. These fall into four main groups: (1) neuro-anatomical methods, (2) electrical recording, (3) stimulation methods, and (4) hormonal and biochemical methods.

Neuroanatomical methods usually consist first of some means of destroying some particular part of the nervous system and then of determining exactly what was destroyed so that the loss of a structure may be correlated with some behavioral variable. Destruction of a neural structure results in a *lesion* or an *ablation*, lesion being the more general and frequently employed term. Two methods of making lesions are in general use: (1) The classical method is one of open, or gross, surgery, in which the brain is exposed and the lesion is made by cutting or burning the structure of interest; (2) a newer and, for many purposes, more precise method that is particularly suited for work with deep structures is the electrolytic method. In this case the lesion is made by passing current through the tip of an electrode located in the structure of interest. To determine afterward what lesion has been made by one or the other of these methods, the brain is usually sliced, mounted on slides, and stained to show cells that have been destroyed or left intact. Because injury to any part of a neuron in the central nervous system causes destruction of the whole neuron, the method of degeneration, that is, of tracing the pathways and centers of dead cells, is often used in reconstructing the lesions that have been made.

A second major technique used in physiological psychology is that of electrical recording of potential changes in various parts of the nervous system. The galvanic skin response (GSR) is one such recording made on peripheral nervous effects. Recordings made from the central nervous system fall into three main categories: (1) an electroencephalogram (EEG) of the spontaneous fluctuations in potential ("brain waves"), (2) gross action potentials of nerves or of sensory pathways, and (3) microelectrode recordings made from single cells or small groups of cells. The last two types of recording are usually "time locked," that is, made with reference to some sensory or electrical stimulus applied at a given time for a given duration. [See NERVOUS SYSTEM, article on ELECTROENCEPHALOGRAPHY.]

Stimulation methods, a third class of methods, consist of the electrical or chemical stimulation of the brain. In the past, electrical methods have been most prevalent, but recently techniques for chem-



ical stimulation have been improved and are coming into greater use. Either method may be used in exploratory or chronic form. In exploratory form the stimulus is applied at different points in the exposed brain, and the effects are noted. The electrical-stimulation method was first used to map the motor cortex; in recent years it has sometimes been used with conscious human patients to map motor, sensory, and association areas of the cerebral cortex. Chronic stimulation may be achieved by setting in place surgically an implanted electrode (for electrical stimulation) or cannula (for chemical stimulation) and after surgical recovery, using the implant, at will, to stimulate the otherwise normal, intact animal. Electrical stimulation applied in this way may be either aversive or rewarding. Chemical methods of stimulation through implants are chiefly being used to study the structures involved in such physiological drives as hunger, thirst, and sex. [See NERVOUS SYSTEM, *article on* BRAIN STIMULATION; *see also* DRIVES.]

The fourth general class of physiological methods, namely the hormonal and biochemical methods, consists of any intervention in the biochemical systems of the body. The oldest form of this method is the removal of an endocrine gland, such as the gonads, to determine its effects on behavior. This gland removal may be combined with replacement therapy, in which the hormone normally made by a gland is injected in its absence. Replacement therapy has the advantage of allowing the reversal of hormonal conditions by the experimenter at some later time. In recent years, with advancing knowledge in psychopharmacology and in biochemistry, chemical methods increasingly employ the administration of drugs or of other biochemical substances through the mouth, directly into the blood stream, and sometimes to the brain itself. [See DRUGS, *article on* PSYCHOPHARMACOLOGY; SEXUAL BEHAVIOR, *article on* SEXUAL DEVIATION: PSYCHOLOGICAL ASPECTS.]

### Major areas

Physiological psychology concerns the same areas as human and animal experimental psychology: (1) sensory processes, (2) motor functions, (3) motivation, (4) learning, and (5) complex processes, including language, thinking, and social behavior. The amount of information available about each area is progressively smaller in the order mentioned. This comes about partly because techniques for the psychological study of sensory and motor functions are more highly developed and partly because more complex processes are most prominent in higher animals and man, where

it is difficult or sometimes impossible to employ the appropriate physiological techniques.

The remainder of this article will summarize the principal conclusions that can be drawn from the field of physiological psychology at the present time. The purpose is to give an over-all picture that will integrate and supplement the several articles in this encyclopedia covering or touching upon specific topics in physiological psychology.

**Sensory systems.** Elsewhere in this encyclopedia are separate articles covering the various senses. They therefore will not be discussed separately here. Rather we shall be concerned with the general features of sensory function.

Each of the senses responds to a different form of energy: electromagnetic, mechanical, chemical, etc. Whatever the energy of stimulation, however, the first step in the processing of sensory information is a transduction of energy into an electrical form known as the generator potential. This is a graded, fluctuating potential that does *not* follow the well-known all-or-none law. Rather its magnitude is proportional to the intensity of the impinging stimulus. When the stimulus is a vibration or oscillation, as in hearing, the generator potential also oscillates and reflects rather faithfully the transitory changes in the stimulus. When the stimulus is a steady one, as in continuous stimulation of the eye by light, the generator potential also rises to, and maintains, some steady value.

The transduction process in which the generator potential arises is different in the different senses. In the eye a photochemical event intervenes; light decomposes photosensitive molecules, and this decomposition gives rise to generator potentials. In the ear, pressure changes in the inner ear act on hair cells, which generate a potential. Other senses have not been so carefully studied in this regard, but in each case the intervening event giving rise to the generator potential seems to be either mechanical or chemical.

The generator potential in turn gives rise to the now-familiar nervous impulse. This follows the all-or-none law of either not occurring at all or of appearing at a fixed amplitude. These nervous impulses, or in physiological parlance, spike potentials, propagate inward from the sense organ to the central nervous system and are the sensory messengers. Their number is proportional to the size of the generator potential and hence reflects the intensity of the stimulus.

The identity of different kinds of receptors is maintained fairly well in the sensory pathways but not in any simple fashion. In the eye, for example, there are at least three different kinds of color

receptors, each most responsive to a different part of the visible spectrum. Farther along the visual pathways, in the thalamus and cerebral cortex, one can find, using microelectrodes, nerve cells that respond specifically to different parts of the spectrum. At this level, however, there is evidence of an interaction of receptors. Different receptors play upon the same central neuron, some exciting and some inhibiting it, and its response therefore is a complex function of the activities of different receptors. Similar statements can be made for all the senses. Many central neurons respond specifically to certain patterns of stimulation on the periphery.

All sensory systems have a topographical projection. That is to say, they maintain in various degrees the spatial relations of the peripheral receptors. The most precise case of topographical projection is found in the visual system, where there is a point-to-point projection of the retina on the visual cortex. Such a projection distorts the peripheral map, a relatively larger area of the cortex being given over to the highly acute fovea than to the less acute peripheral retina. This kind of distortion is seen in all the senses, but topographical relations are nevertheless maintained.

Generally speaking, spatial and qualitative aspects of sensory experience are mediated at the cortical level, and the intensive aspects of experience are processed subcortically. In higher vertebrates, for example, removal of the visual cortex causes pattern blindness but leaves the animal able to discriminate differences in intensity. Similarly, animals deprived of auditory cortex are deficient in sound localization and in pitch discrimination but are able to respond to the presence or absence of sounds if they are not too weak. Similar examples could be given in the other senses.

To summarize, then, peripheral stimuli impinging on receptors are first transduced into graded generator potentials. These give rise to all-or-none spike potentials which propagate centrally. Many central neurons have specific functions representing different kinds of peripheral stimulation, but they reflect interaction of many neurons playing upon them. All sensory systems maintain a topographical orientation. In general, the cerebral cortex is necessary for responding to the spatial and qualitative aspects of stimulation but not to its intensive aspects. [See NERVOUS SYSTEM, article on STRUCTURE AND FUNCTION OF THE BRAIN.]

**Motor functions.** Let us turn now to the motor functions. These may be considered, for the present purposes, in three general classes: (1) skilled acts, (2) postural reactions, and (3) emotional reac-

tions. In higher vertebrates the ability to make highly skilled motions, such as the manipulation of objects, depends on the integrity of the cerebral cortex and even more specifically on the motor (precentral) area of the cortex lying just in front of the central fissure. Removal of this area, including some fringe portions of it lying behind the central fissure, causes paralysis.

The maintenance of tone in skeletal muscles, the keeping of one's balance and upright posture, and certain other specific reactions, such as righting reflexes, are included in the term *postural reactions*. These depend on a complex system in the brain which includes several parts of the cerebral cortex, principally those lying in front of and behind the motor cortex, several subcortical structures, and the cerebellum of the hindbrain. There are certain specific effects of injury to any one of these structures, but the system also tends to work as a whole, so that the more of it that is intact the better the postural adjustment of the organism.

Emotional reactions depend upon both cortical and subcortical structures of the cerebrum, but the most important ones are subcortical. The cerebral cortex exercises considerable control over the timing and intensity of emotional reactions, but the primary emotional mechanisms are to be found in a complex of centers and pathways known as the limbic system. Of these, three structures are most prominent: the septal area, the amygdala, and the hypothalamus. The hypothalamus seems to be the principal organ of integrating patterns of emotional expression, but it is controlled or modulated by opposing influences from the septum and amygdala. In general, the septum is inhibitory; removing it causes an animal to be hyperemotional. The amygdala, on the other hand, is mostly, though probably not entirely, excitatory; removing it causes docility. [See EMOTION.]

**Motivation.** Turning now to a third area of physiological psychology, that of motivation, we may note that rapid progress has been made in this area in recent years, largely through the use of electrodes and cannulas implanted in the deep structures of the brain. The general picture emerging from this work is as follows: Physiological drives such as hunger, thirst, and sex are based in chemical conditions of the blood. What these conditions are is not always precisely known, but whatever they are, they come to affect behavior by circulating in the blood to the hypothalamus and there affecting certain specific cells and centers. There are, for example, certain cells in the hypothalamus that respond to the female hormone, estrogen, and are responsible for the behavior of, say, a cat in



heat. This is demonstrated by bringing animals into heat by applying small quantities of estrogen to the cells of the hypothalamus and then, by carbon 14 tagging of the estrogen, determining that the estrogen has been taken up by certain cells and not by others. Although not all experiments are as dramatic, many lead to the conclusion that motivated behavior is aroused by certain hypothalamic cells responding much like chemical receptors to chemical conditions in the blood.

Within the hypothalamus, motivational centers appear to work in pairs. This statement is well substantiated for hunger and sleep but is not yet so clear for thirst and sex. In hunger a satiety center is found in the ventromedial region of the thalamus, and a hunger center in the ventrolateral region. One is called a satiety center because its removal causes an enormous increase in food intake (hyperphagia); the other, a hunger center, because its removal causes the opposite effect, namely, such a remarkable loss of hunger that animals starve to death unless artificial means of keeping them alive are employed. A thirst center is difficult to distinguish anatomically from the hunger center, for damage to the hunger center causes a primary disturbance of thirst (adipsia), but there are probably different cells in this region subserving hunger and thirst. Substantial increases and decreases in sexual behavior may follow experimental lesions in the hypothalamus, but the arrangement of hypothalamic centers for sexual behavior has not been clearly defined. In any case, fundamental mechanisms for the control of physiological drives are located in the hypothalamus. These are aroused by chemical conditions in the blood, and they are also, of course, affected by inputs from sense organs and from higher cerebral structures, including the cerebral cortex. [See DRIVES; MOTIVATION; SEXUAL BEHAVIOR, *article on ANIMAL SEXUAL BEHAVIOR*.]

**Conditioning and learning.** A fourth major area of physiological psychology is the study of the neural mechanisms of conditioning and learning. This area is conveniently divided into three subareas: (1) conditioning, (2) discriminative learning, and (3) problem solving and maze learning. [See LEARNING, *especially the article on NEUROPHYSIOLOGICAL ASPECTS*.]

Contrary to Pavlov's assumption in his early work on conditioning, the cerebral cortex is not necessary for conditioning to take place. Animals that are completely decorticate can be conditioned in the classical sense to make some response to a signal. Difficulty arises, however, when an attempt is made to produce an instrumental response, such

as raising of the paw to a signal in order to avoid shock. The difficulty is partly that decortication considerably reduces the motor ability of the animal and partly that the decorticate is more easily aroused emotionally by the noxious shock. If care is taken to keep the animal quiet, by presenting a weak shock and allowing a considerable interval between conditioning trials, even an instrumental response may be conditioned.

Although we are certain that the cerebral cortex is not necessary for conditioning, it is hard to say just what parts of the nervous system are. Although some work has been done on spinal animals, it is doubtful that conditioning can take place at the level of the spinal animal. On the other hand, conditioning has been obtained in decerebrate animals whose brains have been transected at the midbrain level. Hence, it appears that the simple phenomenon of conditioning can occur at the level of the hindbrain.

In recent years there has been considerable interest in the electrical correlates of conditioning. This interest has been pursued by recording the electroencephalogram or other oscillations in brain activity at different sites in the brain during the course of conditioning. It is difficult as yet to draw many conclusions from this work. One result, however, does turn up repeatedly. In the reticular activating system, as well as in other structures related to it, an increase in electrical activity accompanies the early stages of conditioning. As the conditioned response is perfected, however, this activity usually subsides. This electrical correlate appears to reflect attention and arousal. As conditioning trials proceed, the animal becomes more alert and attentive to the conditioning situation. But later, as the conditioned response becomes more "automatic," such attention wanes and along with it its electrical correlate. [See LEARNING, *articles on CLASSICAL CONDITIONING, INSTRUMENTAL LEARNING, and AVOIDANCE LEARNING; see also ATTENTION*.]

**Discriminative learning** is learning in which the organism learns to make one response to one stimulus and another response (sometimes this is the withholding of response) to another stimulus. Most of the results of research on the neural mechanisms of discriminative learning can be predicted from a knowledge of the sensory mechanisms discussed above. Each sensory system has its area of projection in the cerebral cortex. This projection is concerned primarily with spatial and qualitative aspects of a stimulus. Hence, an animal lacking its visual cortex is unable to learn a visual-pattern discrimination, and one deprived of its auditory cortex is unable to make good sound localizations.

In each case, however, the deficit appears to be one of sensory capacity rather than an inability to learn apart from sensory capacity.

The problem becomes more intricate, however, when we turn to intensity discrimination. The capacity to make such a discrimination is not dependent upon the cerebral cortex. Yet animals that have learned a discrimination and are then subjected to removal of the relevant projection area usually lose all memory of the discrimination they have learned. On the other hand, however, they are capable of relearning it in about the same number of trials as they required originally. It is this phenomenon, which has been obtained many times, that led Lashley to propose the principle of equipotentiality, namely, that when one part of a sensory system is injured, other parts of the system have the capacity to take over the function impaired by the injury. This, however, is not a very satisfying explanation, and the matter is still under study. [See *LEARNING, article on DISCRIMINATION LEARNING.*]

Problem solving and maze learning are in the class that Edward L. Thorndike called trial-and-error learning. Put in a problem box, the animal tries one thing, then another, and finally strikes on the "trick"—which might consist of operating a latch—that solves the problem, say, of obtaining food. Or put in a maze, the subject wanders in and out of blind and correct alleys until it finally comes to the reward.

In general, the ability to learn by trial and error depends on the cerebral cortex. If the problem is very simple, requiring only some gross movement like stepping on a pedal, most of the cortex can be removed without impairing this learning ability. The more complex the problem, however, the more the cerebral cortex is required. In fairly difficult problems, whose solution borders on requiring the use of reasoning, very small cortical lesions may seriously impair problem-solving ability. However, in all cases a distinction must be made between learning and retention. A lesion made after a problem has been learned may seriously impair retention, but if more training trials are given, the subject may learn the solution again. This is another instance of what Lashley called equipotentiality. Also, for fairly complex problems there is not much localization of function; the mass of cortex intact rather than a particular locus is what determines degree of retention or of learning ability. This is what Lashley called mass action.

In making these statements, we must qualify them according to the species we are talking about. The primate brain is more highly organized than

the rat's brain, and there is a correspondingly higher degree of localization in it than in the rat's brain. In the primate brain, for example, there is a well-defined area of the temporal lobe that is concerned with the learning and retention of visual discriminations. If this area exists in the rat brain, it is not well delineated.

**Language and complex processes.** A fifth major area of physiological psychology is the study of complex processes, such as language and thinking, and their neural correlates. Although the most interesting of all the areas of physiological psychology, it is the one about which the least is known. This is partly because of the difficulty of studying brain functions in man, in whom these processes are most highly developed, and partly because the processes are inherently difficult to study with any precision. There are, however, certain conclusions that can be drawn at this time. One is that there is generally more localization of function in man and primates than in lower animals, although this localization is by no means precise. More specifically, we find that the prefrontal lobes lying in front of the motor areas are concerned with expressive functions, particularly with those that require planning and doing things in a prescribed order. The posterior cortex is more concerned with receptive functions. In it are the primary sensory projection areas found in lower animals, but also quite massive "association" areas. Many of these areas have either an unknown or a diffuse function, but some are clearly concerned in certain kinds of memory and recognition abilities.

Language, of course, holds a peculiar fascination because it is the most complex of all behavior functions and is uniquely human. For just that reason, however, our knowledge of its physiological mechanisms is limited, being based solely on clinical cases of brain injury. It is fairly well established, however, that the left side of the cerebral cortex is almost always the seat of language functions. Injuries to the right cerebral cortex seldom affect language ability, while those to the left, if of any consequence, almost always do. A large area of the left hemisphere is concerned in language; it extends forward into the frontal lobes and backward into the temporal and parietal lobes. Within this large area there is some localization of different kinds of language functions, but not a great deal. As research proceeds and our methods of analysis become more refined, we should gain a clearer picture of the way the brain coordinates complex functions. [See *LANGUAGE and MENTAL DISORDERS, article on ORGANIC ASPECTS.*]

CLIFFORD T. MORGAN



[Directly related are the entries DRUGS, article on PSYCHOPHARMACOLOGY; MENTAL DISORDERS, article on BIOLOGICAL ASPECTS; NERVOUS SYSTEM; SENSES. Other relevant material may be found in HEARING; PAIN; SKIN SENSES AND KINESTHESIS; TASTE AND SMELL; VISION; and in the biographies of BROCA; FLOURENS; LASHLEY.]

## BIBLIOGRAPHY

- COUNCIL FOR INTERNATIONAL ORGANIZATIONS OF MEDICAL SCIENCES 1961 *Brain Mechanisms and Learning*. Edited by R. W. Gerard, A. Fessard, and J. Konorski. Springfield, Ill.: Thomas.
- DETHIER, V. G.; and STELLAR, ELIOT (1961) 1964 *Animal Behavior: Its Evolutionary and Neurological Basis*. 2d ed. Englewood Cliffs, N.J.: Prentice-Hall.
- Handbook of Physiology. Section 1: Neurophysiology. Edited by John Field et al. 3 vols. 1959-1960. Baltimore: Williams & Wilkins.
- ISAACSON, ROBERT L. (editor) 1964 *Basic Readings in Neuropsychology*. New York: Harper.
- LASHLEY, KARL S. (1915-1958) 1960 *The Neuropsychology of Lashley: Selected Papers*. Edited by Frank A. Beach et al. New York: McGraw-Hill. → Published posthumously.
- MORGAN, CLIFFORD T. (1943) 1965 *Physiological Psychology*. 3d ed. New York: McGraw-Hill.
- POLIDORA, V. J.; and BOWMAN, R. E. 1966 *Psychochemistry*. Pages 183-208 in Joseph E. Sidowski (editor), *Experimental Methods and Instrumentation in Psychology*. New York: McGraw-Hill.
- SHEER, DANIEL E. (editor) 1961 *Electrical Stimulation of the Brain: An Interdisciplinary Survey of Neurobehavioral Integrative Systems*. Austin: Univ. of Texas Press.
- SYMPOSIUM ON THE FRONTAL GRANULAR CORTEX AND BEHAVIOR, PENNSYLVANIA STATE UNIVERSITY, 1962 1964 *The Frontal Granular Cortex and Behavior*. Edited by J. M. Warren and K. Akert. New York: McGraw-Hill.
- SYMPOSIUM ON INTERDISCIPLINARY RESEARCH, UNIVERSITY OF WISCONSIN, 1955 1958 *Biological and Biochemical Bases of Behavior*. Edited by Harry F. Harlow and Clinton N. Woolsey. Madison: Univ. of Wisconsin Press.
- TEITELBAUM, PHILIP 1967 *Physiological Psychology*. Englewood Cliffs, N.J.: Prentice-Hall.
- UHR, LEONARD; and MILLER, JAMES G. (editors) 1960 *Drugs and Behavior*. New York: Wiley.
- WOOLDRIDGE, DEAN E. 1963 *The Machinery of the Brain*. New York: McGraw-Hill.
- ZANGWILL, O. L. 1960 *Cerebral Dominance and Its Relation to Psychological Function*. Edinburgh: Oliver.

## III

## CONSTITUTIONAL PSYCHOLOGY

The term *human constitution* refers to the sum total of the morphological, psychological, physiological, and biochemical characteristics of an individual which are mainly determined by genetic factors but which are influenced in varying degrees by environmental conditions. One of the main objectives of the study of the human constitution is the discovery of significant relationships among

these characteristics and between them and pathological and behavioral characteristics. The field of study includes ascertaining the nature of individual variations, their determinants, and the extent to which they are governed by genetic and environmental influences respectively. Those attributes that show marked variability between individuals and relative constancy within the individual are the ones most likely to be useful for research purposes.

**Historical perspective.** Throughout the ages there has been interest in finding clues to a person's personality, temperament, behavior, and susceptibility to disease. Aristotle, for example, founded the science of physiognomy, which attempted to relate a person's mental and moral qualities to his external lineaments. Hippocrates and Galen described specific physical types with characteristic temperamental dispositions and specific susceptibility to different diseases.

In more recent times, Beneke (1881) published data on organ size in various diseases and in relation to various physical types. During the same period, Giovanni, the founder of clinical anthropometry, tried to establish that the cause of the special morbidity of an organ was based on its disproportionate development relative to other parts of the body (1891). Kretschmer (1921) published his classic study *Physique and Character*, in which he described relationships between different types of physique and predisposition to different psychiatric illnesses. He considered that there is a clear biological affinity between manic-depressive illness and the broad or pyknic type of body build and that schizophrenic patients commonly exhibit the characteristic features of the narrow or leptosomatic physique and certain types of dysplasia, or physiques in which there is uneven growth in different body regions. Kretschmer also extended the scope of his theory to the relationship between physique and normal personality types. He considered that there is a gradation of temperament characteristics from the manic-depressive through the cycloid psychopathological form to the normal cyclothymic personality and, likewise, a gradation of characteristics from schizophrenia through schizoid psychopathology to the normal, well-adjusted schizothyme. He considered that schizoids and schizothymes tend to have leptosomatic, athletic, and dysplastic physiques and that cycloids and cyclothymes tend to have pyknic physiques.

Kretschmer's work stimulated a great deal of research and, in fact, was responsible for the recrudescence of interest in constitutional psychology, which had, up until then, been relegated to the

background by the growth of knowledge in pathology and bacteriology relating to the causation of disease. [See the biography of KRETSCHMER.]

### Major approaches

In recent years there have been considerable advances in the study of constitutional psychology through the development of new techniques and methods of investigation. The following are the main approaches:

Factor analysis and other statistical procedures for elucidating the variables underlying body build; Sheldon's classification of somatotypes; Lindgard's method of assessing body build; analysis of tissue components; studies of androgyny; and genetic and growth studies.

From the time that Hippocrates divided physiques into the *habitus phthisicus*, or narrow type, and the *habitus apoplecticus*, or broad type, many different systems of classifying physiques have appeared. These are summarized by Rees (1960), who noted that the majority of authors had independently described more or less corresponding physical types with a dichotomy of contrasting, antithetical types as extremes, and with intermediate types in some classifications. The basis of these classifications was not scientific but reflected preconceived ideas about the nature of variations of physique and suppositions based on alleged physiological, dietetic, and developmental characteristics.

The concept of physical types as disparate or mutually exclusive categories has now been replaced by the concept of a continuous variation in body build. The field of constitutional research had been handicapped by the fact that the full extent of the variations in physique in the normal population was not made known and that the various proposed typological classifications were based on unverified hypotheses. What was needed was an inductive method that would ascertain the nature of the variations of physique in the normal population and describe these variations most effectively and economically. Factor analysis, which had proved fruitful in the study of intelligence, provided one such method.

**Factor analysis.** Factor analysis provided a relatively objective and inductive method of elucidating the variation of physique in the normal population and describing it in an economical and effective way. Factor analysis reduces a large number of variables to a smaller number of categories or factors that can account for the correlations between the original variables. It has a special advantage for the study of the human constitution in that it can account for a representation of human

physique which is far more comprehensive than that given by anthropometric measurements alone. [See FACTOR ANALYSIS.]

The method of factor analysis may be applied to traits or to persons. The factor analysis of persons provides a method of determining how true an individual is to a particular type and involves the specification of the type in question by a set of representative measurements expressed in standard measure. The correlation between the person's measurements and those of a standard pattern indicates how nearly he approximates a perfect representative of that type. This method presupposes that the characteristics of the type have already been clearly established; it is, therefore, not used for ascertaining the nature of the variations in human physique that are useful in delineating physical types.

Most factor analyses of physique deal with traits; the steps required for employing factor analysis on physical constitution in this way are as follows: First, a random and representative sample of the normal population is selected and a comprehensive series of anthropometric measurements covering all parts and dimensions of the body is taken on the sample. Each measurement is correlated with all the other measurements for the total group, providing a correlation matrix. One of the primary assumptions in factor analysis is that when a correlation exists between any two variables, they must share a common factor.

The next step is the discovery of the factors accounting for the correlations. The effect of each discovered factor is removed from the matrix before extracting the next factor. This process is repeated until the residual correlations are statistically insignificant. Some factor techniques may extract factors that are independent, that is, factors that are themselves uncorrelated or orthogonal. Other techniques involve matrix rotation, in an attempt to get a better description of groups of correlations and produce correlated or oblique factors.

Factor analysis also indicates which measurements are likely to be most effective for discriminating physical types. Such measurements may be combined to form indexes of body build, either in the form of suitably constructed ratios or in the form of regression equations of weighted measurements.

To illustrate the use of factor methods in analyzing physique, the results obtained by Rees (1950a) for a group of 200 adult men and a group of 200 adult women are presented briefly. Some 18 anthropometric measurements were taken on both groups and were intercorrelated and factor analyzed. For



both men and women the correlation coefficients between all physical measurements were positive. A general factor was extracted first and was found to contribute 34 per cent of the variance. The second factor extracted was bipolar, having positive saturations with length measurements and negative saturations with breadth and circumference measurements. These two factors were found in both the male and female groups. This second factor clearly differentiates two extremes: the linear type of physique—length measurements preponderant in relation to breadth and circumference; and the broad type—breadth, width, and circumference measurements preponderant over length.

In the males two measurements, namely, stature and transverse chest diameter, had particularly high loadings on the second factor. These measurements were combined as follows to form the Rees-Eysenck index of body build:

$$\frac{\text{Stature} \times 100}{\text{Transverse chest diameter} \times 6}$$

In the female group there were four measurements with high loadings on the second factor: stature, height of the top of the symphysis pubis from the ground, hip circumference, and chest circumference. It was not feasible to construct a simple female index along the lines of the one for men; thus, the Rees index of female body build was devised as a regression equation based on standard measurement units (Rees 1950a):

$$0.59 \text{ stature} + 0.47 \text{ symphysis height} \\ + 0.31 \text{ chest circumference} \\ - 0.64 \text{ hip circumference.}$$

These indexes have been calculated for large groups of subjects and in all instances there was a unimodal frequency distribution approximating the normal frequency curve. There was no evidence of multimodality or bimodality as there would be if the physical types represented mutually exclusive, disparate categories. What was found was a continuous variation in physique from one extreme to the other, the extremes of the distribution being present only as easily noticeable but infrequently occurring types.

Any such division into physical types, although convenient and sometimes necessary for research, is arbitrary. One standard deviation above and below the mean were considered to be statistically acceptable arbitrary demarcation points dividing the distribution into three classes: (1) leptomorphs, with a preponderance of length dimensions compared with breadth, width, and circumference; (2) eurymorphs, with large circumference, breadth,

and width measurements in comparison to length; (3) mesomorphs, with an intermediate relationship between lateral and circumference dimensions on the one hand and linear measurements on the other. These categories of physique are not intended to convey the impression of disparate types and are, in fact, ranges of body build objectively measured by new indexes of physique and demarcated by statistical criteria for purposes of description and research.

These findings, derived from groups of adult British men and women, are consistent with those of a large number of other researchers using factor methods in groups of different ages, racial composition, and socioeconomic status. Burt (1938), who was the pioneer of the factor analysis of physique, described the results he obtained for groups of British children and adults and found a well-marked general factor, as well as factors accounting for disproportionate growth in length on the one extreme and breadth, thickness, circumference, and weight on the other extreme. A summary of these studies is presented in detail by Rees (1960).

The factors obtained from factor analysis of anthropometric measurements will, of course, depend on the selection and the number of original measurements used. Howells (1952) carried out a factor analysis on physical measurements of university students; the data included a large number of head and face dimensions. There were three limb measurements and three facial-length measurements. Howells extracted a factor for limb and facial length and concluded that the extraction probably depended more on the measurements chosen than on growth patterns. In all, he found seven factors, which he interpreted as relating to general body size, long-bone length, cranial size, brain size, lateral facial cranial development, face length, and ear size. These can be grouped into four size factors, two length factors, and a breadth factor.

Sheldon's varieties of physique. Sheldon's system of classifying physique (1940) was based initially on the inspection of standardized photographs of a large number of individuals. He concluded that there are three components in physique: endomorphy, mesomorphy, and ectomorphy. These, he considered, consist of tissues derived from the three germinal layers: the endoderm, the mesoderm, and the ectoderm. Each subject was observed and rated on three 7-point scales, one for each component. For example, a rating of 7-1-1 represents extreme endomorphy, preponderant in breadth and circumference measurements; a 1-7-1 rating

represents extreme mesomorphy, with marked bone and muscle developments, broad shoulders, and muscular limbs; a rating of 1-1-7 reflects extreme ectomorphy, characterized by linearity and relatively poor development of breadth, width, and circumference measurements.

Sheldon's method was an advance over previous typologies, which tended to regard physical types as disparate and mutually exclusive categories. Sheldon's method has been criticized because of its subjectivity and because of evidence showing that an individual somatotype, contrary to Sheldon's claim, is not necessarily constant for a particular individual. Changes in a person's somatotype occur with changes in his diet and with increase in his age.

In spite of its subjectivity, it has been demonstrated that retest reliability of ratings by experienced workers is high. Sheldon's assumption that his three components are composed of tissues derived from the three primary germinal layers has not been established, and the number of components underlying variations of physique described by Sheldon is now under considerable doubt because a number of workers have demonstrated that the data can be fully explained in terms of two components rather than three.

Humphreys (1957) found that the correlations between endomorphy, mesomorphy, and ectomorphy provided by Sheldon indicate the existence of only two independent (but not necessarily valid) types of physique. Furthermore, Howells (1952), using factor analysis of persons who were markedly dominant in one or another of the three components, found that the factors extracted do not correspond with Sheldon's three components. He found that ectomorphy and endomorphy are not independent factors but are opposite extremes of a continuum and manifestations of a single factor.

*Modifications of Sheldon's method.* Parnell (1958) pointed out some inherent disadvantages in Sheldon's system, namely, the subjectivity of the method of estimating somatotype dominance, personal objections by subjects to being photographed in the nude, and technical difficulties in obtaining the equipment necessary for carrying out photography for somatotyping. Parnell's method involves taking measurements of height, weight, bones, muscle, girth, and skinfolds of subcutaneous fat. Parnell uses ratings of fat, muscularity, and linearity in place of Sheldon's endomorphy, mesomorphy, and ectomorphy (bone development is included in the estimate of muscularity), and an individual is rated according to these three factors. Primary dominance is indi-

cated by the capital letters *F*, *M*, *L* and secondary dominance by the small letters *f*, *m*, and *l*. The disadvantage of this system is the use of fat and muscularity, which can show relatively marked variation in the individual during his lifetime, sometimes within a comparatively short period of time.

*Lindegard's assessment of body build.* Lindegard (1953) concluded from his studies that there are four factors responsible for the variations in body build: length, sturdiness, muscle, and fat. His method involves rating a person's dimensions on scales representing each of these factors. The length factor is derived from the length of the radius and tibia; the sturdiness factor is derived from the breadth of the femoral condyle, bimalleolar and bistyloid breadths, and the distance between the nasion and the sella turcica, measured by X-ray cephalometry; the muscle factor is derived from dynamometer readings of hand grip and foot press; and the fat factor is derived from girth and body weight and girth of the extremities.

The length, sturdiness, and muscle factors are determined directly, but the fat factor is deduced by comparing the weight and extremity girths with the skeletal build and muscle factor. The four factors are plotted on a graph in terms of standard deviation.

Lindegard found that length, sturdiness, and muscle factors are interdependent, and he dealt with this by regression techniques, which yield a relative sturdiness factor and a relative muscle factor. The relative sturdiness factor expresses the skeletal sturdiness in relation to the length factor. The relative muscle factor is expressed in terms of the dynamometer readings in relationship to stature.

The weakness of Lindegard's system is the probable inconstancy of dynamometric recordings in an individual and the possibility that they might be altered within a short space of time by training and exercise and are influenced by age, disease, etc.

*Tissue components, androgyny, gynandromorphy.* The measurement of the relative development of bone, connective tissue, fat, and muscle is a comparatively recent development and could be a useful complementary method to factorial studies of physique and to Lindegard's system and Sheldon's somatotyping method. The usual method is to use standardized X rays of the calf to measure the relative thickness of bone, fat, and muscle. The relative amounts of these factors in the calf have been found to be more or less independent of each other.

The terms *androgyny* and *gynandromorphy* refer respectively to the degree of femininity of body



build in the male and the degree of masculinity of body build in the female. Two distinct methods have been employed to assess androgyny: subjective procedures, consisting of assessment and rating of general appearance and special features, and objective anthropometric measurements, such as those described by Tanner (1951).

Tanner used the method of discriminant analysis to ascertain the best possible discrimination between males and females with regard to biacromial diameter and bi-iliac diameter. He found that the best index for discriminating the sexes is

$$\text{Biacromial diameter} \times 3 - \text{Bi-iliac diameter} - 82.$$

Tanner (1955) considers that androgyny, in contrast to somatotype, is determined by events taking place at adolescence. Androgyny scores have been used recently to study relationships with academic and athletic performance, physical efficiency, and various diseases.

**Genetic factors and growth processes.** Important morphological characteristics such as body size, body shape, and body composition are a function of growth rates, which, in turn, are determined by genes in interaction with environmental influences. Polymorphism, or natural variability, is a consequence of genetic and physiological reaction to the environment.

Twin studies are useful in ascertaining the relative importance of genetic and environmental factors in determining various bodily dimensions. Dahlberg (1926) compared the variability in anthropometric measurements between nonidentical and identical twins. It was found that 90 per cent of the variability of body length measurements, and slightly less for body breadth, head, and face dimensions, was determined by heredity.

The majority of bodily dimensions, as well as variations in body size and body shape, are distributed along a normal frequency curve. This is in keeping with the hypothesis of multifactorial inheritance of these characteristics.

It has been demonstrated in animals that general body size is largely determined by genetic factors. It is particularly interesting to note that Gregory and Castle (1931) demonstrated that differences in growth, measured by the number of blastomeres as early as 36 hours after fertilization, exist between large and small varieties of rabbits.

Skeletal age, measured by the degree to which ossification centers and epiphyses are developed, has been shown by Reynolds (1943) to be largely genetically determined, with a correlation of .71 in identical twins, a correlation of .28 in siblings, .1 in first cousins, and .01 in unrelated persons.

Age at menarche is a convenient measure of the rate of growth. According to Tanner (1960), the age at menarche is closely related to the growth spurt in body measurements and to the maturation of the skeleton. The age at menarche is normally distributed, which suggests that menarcheal age depends upon the combined action of genes at many different loci rather than any single allele. Tanner reports that in identical twins the average difference in months of age at menarche is 2.8, in non-identical twins 12, in sisters 12.9, and in unrelated girls 18.6.

It has been known for many years that girls are physically more mature than boys at all ages from birth until adulthood. The recent discovery of some persons with chromosomal formulas XXY or XO has provided an opportunity of studying genetic factors that influence differential sex growth rates. Persons with an XXY chromosome complement suffer from Klinefelter's syndrome and in appearance are males but have small, atrophic testicles, abnormally long legs, and narrow shoulders. Those with XO chromosomes suffer from Turner's syndrome and are apparently females but lack ovaries, are short in stature, and have no development of the secondary sex characteristics at puberty. [See *SEXUAL BEHAVIOR, article on SEXUAL DEVIANCE: PSYCHOLOGICAL ASPECTS.*]

The question arises whether the skeletal maturity of an XXY child would follow that of the male or the female. Tanner (1960) reports an investigation into the ratio of skeletal age to chronological age in children with Klinefelter's syndrome compared with normal boys. It was found that the ratios failed to differ significantly, but if they were compared with girls, great differences occurred. Thus, XXY follows the course of XY. Similarly, 16 persons with Turner's syndrome were rated and found to follow the normal pattern for girls. Thus, XY and XXY have the male initial retardation of skeletal development, XO and XX do not. The inference is clear that this retardation in males is due to genes located on the Y chromosome.

It is clear that individual differences in adult physique must come about through differential local and general growth rates during childhood.

Tanner (1960) points out that there is much evidence to show that the child's measurements at birth reflect almost entirely the action of the uterine environment and are thus dependent on maternal genotype and maternal environmental factors only. The genotype of the child is practically without effect; it is only subsequently that the child's genes exert their action and become a basis for predicting the child's adult status and that the

child's resemblance to his genetic parents becomes increasingly marked.

### Personality and physique

It will be convenient to consider studies relating personality and physique according to the particular system used for classifying body build.

**Kretschmer's typology.** Kibler (1925), utilizing questionnaires, was able to confirm an association between leptosomatic body build and schizothymic temperament, on the one hand, and pyknic body build and cyclothymic temperament, on the other. However, Pillsbury (1939), using similar methods, found no evidence to support Kretschmer's theory.

A variety of psychological attributes have also been studied in relationship to Kretschmer's physical types. Enke (1928), for example, found that persons of pyknic physique were more easily distracted in reaction-time experiments and were poorer in tachistoscopic perception of nonsense syllables than were those of leptosomatic physique. Pyknics tended to respond predominantly to color in tachistoscopic experiments and also tended to react to colors on the Rorschach test.

Lindberg (1938) found that pyknic body build was associated with a greater tendency to color reaction and leptosomatic body build with a greater reaction to form. Keehn (1953), however, has thrown doubt on the interpretation of the tests used as measures of color-form reactivity. According to Misiak and Pickford (1944), leptosomes are more persevering at tasks and have a lower tendency to perseverate responses than pyknics. Smith and Boyarsky (1943) found that leptosomes had a higher tempo, as measured by their tapping rate, and quicker reaction time than pyknics. In a variety of tests, other workers (such as Klineberg, Asch, & Block 1934) found no significant differences between leptosomes and pyknics.

**Factor analyses.** Burt (1938), in a study of children, found correlations between stoutness of build and cheerful emotions and between thinness of body build and inhibited and depressive tendencies. Rees and Eysenck (1945) and Rees (1950a; 1950b; 1950c), in studies of adults of both sexes, found that leptomorphs have a greater tendency to introversion and a higher tendency to somatic manifestations of emotional tension and autonomic dysfunction, whereas eurymorphs had a greater tendency to extroversion and a greater tendency to hysterical symptoms.

Sanford and his associates (1943) obtained similar results, finding the tall, narrow type of physique to be associated with autonomic dysfunction and manifestations indicative of dysthymia.

The wide, heavy type correlated negatively with autonomic dysfunction and positively with secure feelings and lively self-expression.

**Sheldon's somatotypes.** Sheldon (1942) postulated three components of personality: viscerotonia, somatotonia, and cerebrotonia. Viscerotonia is characterized by relative love of comfort, pleasure in digestion, need for affection and approval, and the need of company when in trouble. Somatotonia is characterized by an assertive posture, need for exercise, directness of manner, unrestrained voice, and a need for action in trouble. Cerebrotonia is shown by restrained posture, quick reaction, inhibited social reactions, vocal restraint, and need of solitude when in trouble.

In a group of 200 university men between 17 and 31 years of age, Sheldon (1942) found high positive correlations between the components of physique and the components of temperament: between endomorphy and viscerotonia, the correlation coefficient was .79; between mesomorphy and somatotonia, .82; and between ectomorphy and cerebrotonia, .83.

Humphreys (1957) criticized Sheldon's methods and stated that the evidence provided by Sheldon supported not more than two independent (but not necessarily valid) types of temperament. Child and Sheldon (1941) reported only low correlations between somatotype and various other psychological attributes. Tanner (1951) found little to support Sheldon's claims for high correlations between somatotype and temperament.

Parnell (1958), in a study of a group of 7-year-old children, found that fat children had fewer handicaps in the first seven years, were able to express feelings easily, and were confident and sociable. Muscular children were relatively free from anxiety shyness. Linear children were oversensitive, shy, and had greater difficulty in communicating their thoughts and feelings to people around them; they had signs of emotional unrest such as nightmares, panics, temper tantrums, and nail biting and also tended to be meticulous and conscientious and to exhibit high standards.

**Studies based on Lindegard's method.** Lindegard and Nyman (1956) studied a group of 320 men whose physiques were rated according to Lindegard's method (1953). It was found, for example, that the muscle factor was correlated with expansivity and self-confidence versus unobtrusiveness; the more muscular person is more expansive and confident, the less muscular person is emotionally cool and sophisticated.

Wretmark (1953) found that stability (identified as emotional warmth, rather than emotional



coldness, together with sophistication and proficiency of motor activity) was the only attribute correlated with body build. He found that sub-stable individuals were relatively wide and heavy and that extraordinarily stable individuals tended to be narrow in body build.

### Physical attributes and intelligence

For many years research has been carried out to discover whether some physical attribute would provide some clue to a person's intellectual ability. This has been a matter of interest to educators and psychologists who aim to provide objective aids for assessing intelligence.

Rees (1960), reviewing other literature as well as his own investigations, concluded that the evidence indicated no relation between absolute anthropometric measurements and intelligence level.

When the relationship between body type or body build and intelligence is considered, Rees (1960) found that there was a slight tendency for leptomorphs to be higher in intelligence than eury-morphs.

Parnell (1958) investigated the intelligence level of a group of 7-year-old children and found that children with linear physiques were best in reading ability. At the age of 11, linear children did better than general expectation on the entrance examination for British grammar schools. Studies by Parnell (1958) of university students again provided evidence that linear men tend to do better. On the other hand, the best order of performance at work and at games was achieved by *Mf* men, that is, those in whom muscle is the dominant factor and fat the secondary factor.

Rees's conclusion (1960) still holds: the correlation between physique and intelligence is so low that a knowledge of a person's physical status will be of no value in the prediction of intelligence in the individual case.

**Body size.** The Rees index of body size, represented by the product of stature and transverse chest diameter, was obtained for a sample of 1,100 military personnel admitted to the Mill Hill Emergency Neurosis Centre; the frequency distribution was normal and was arbitrarily divided at points one standard deviation above and below the mean into three groups: microsomatics, mesosomatics, and macrosomatics.

A detailed analysis of psychological attributes in these groups showed, for example, that neurotics with large body size tend to be more active and to have a more stable personality, better musculature, a higher intellectual level, and a better prognosis

for return to military duties. The microsomatic group tend to be of inadequate personality make-up: i.e., they have a higher incidence of weak, dependent, anxious, and hypochondriacal traits of personality and narrow interests and hobbies.

It is interesting to note that the psychological correlates of body size differ from those associated with physical type. Whereas the extremes of body type, the eury-morph and the leptomorph, are associated respectively with hysterical and anxiety reactions, there were no significant relations between these reactions and body size.

**Body disproportion.** Seltzer (1946) paid particular attention to disproportionate development of the body, which he defined as values of the index derived from anthropometric measurements which are markedly divergent from the mean, usually in one direction. One such disproportion might be a combination of wide shoulders and thin legs; another, broad hips relative to shoulder width. The term "disproportion" must not be confused with the term "dysplasia." It does not carry any stigma of abnormality; it represents deviations within the normal range of variation.

Seltzer found that individuals, and particularly linear individuals, with unstable autonomic functioning showed a greater tendency to bodily disproportion. He found excesses of one or more bodily disproportion to be associated with the following characteristics: unstable autonomic functions, mood fluctuations, personal instability, inability to adjust, sensitivity, meticulousness, self-consciousness, and introspection. There were significant deficiencies associated with those traits indicative of stability and ease of making adjustment. It was found that disproportions were commoner in marked ectomorphs and least common in marked endomorphs.

**Androgyny.** Using Turner's androgyny scale, Rees (1960) found that a group of 200 soldiers suffering from effort syndrome (neurocirculatory asthenia), a disorder in which breathlessness and various autonomic symptoms accompany exercise, showed more marked feminine attributes in body build than normal, healthy soldiers.

Coppen (1958), using the same index, found that patients who developed toxemia in pregnancy had a higher incidence of masculine features in body build than those who did not.

Parnell (1954) compared the relationship between the relative incidence of masculine and feminine physical traits and academic and athletic performance in men and found that academic distinction was associated with a significant trend

toward feminine traits and that athletes, as would be expected, showed a higher incidence of strong masculine traits.

Lindgard (1956), using a scale based on a regression equation using biacromial and bi-iliac measurements, found no significant relationships with personality variables.

Seltzer and Brouha (1943) found that excellent degrees of physical fitness could be achieved only by persons with a strong masculine component in their physique and that individuals weak in masculinity tended to have poor effort tolerance, that is, the ability to undertake effort without suffering unduly from distress.

### Physique and mental illness

**Neurosis.** Rees and Eysenck (1945) studied a group of 400 soldiers admitted to Mill Hill during World War II. Rees extended this study to 1,000 soldiers (1960) and to 400 women serving in the armed forces who developed neurotic illness (1950a; 1950b; 1950c).

Using the Rees-Eysenck index of body build, it was found that leptomorphic individuals of both sexes tend to have dysthymic symptoms, whereas eurymorphs tend toward hysterical characteristics and symptoms.

Parnell (1958) in an analysis of a survey of 2,000 hospitalized psychiatric patients suggests that if breakdown occurs in *Lf* types it more often occurs before the age of 25. Male patients under age 25 with a diagnosis of depression, anxiety state, psychopathy, or suicidal tendency were found to be generally of the *Lf* physique. Among schizophrenics under age 25, the males tended to be most commonly of *Lf* type, and next most commonly *Lm*. Among female schizophrenics the most frequent type was *Fl*. *Lm* types of both sexes are more disposed to breakdown from 25 to 35 years of age. *Ml* types of both sexes are more predisposed to breakdown from 35 years of age onward. Paranoid schizophrenia and paranoid illnesses are their commonest examples of thought disorders. Depression is the major affective illness in both sexes.

*Mf* types were the most stable in both sexes. When they become mentally ill they are usually over 30 years of age. Thought disorder usually takes the form of a paranoid illness, but an affective disorder with depression is the commoner form of illness. [See NEUROSIS.]

**Manic-depressive illness.** In a review of the literature and his own studies, Rees (1960) concluded that there is some correlation between pyknic habitus and manic-depressive psychosis, but

the correlation is not, by any means, as great as suggested by Kretschmer. Rees (1960) also pointed out that the age factor accounts for part of this correlation but does not completely invalidate Kretschmer's theory. [See DEPRESSIVE DISORDERS.]

**Schizophrenia.** Rees (1943) carried out a detailed anthropometric investigation on a sample of schizophrenics as well as on a control group of normal men of similar age distribution. He found that schizophrenic patients are significantly smaller in stature, chest circumference, hip circumference, biacromial diameter, supersternal height, and symphysis height. Body build indexes showed that they are significantly more leptosomatic than the normal group.

There is some evidence of a relation between type of physique and specific form of schizophrenia. Connolly (1939) reported that paranoid schizophrenics are of larger body size and more pyknic in physique than other types of schizophrenic patients. Rees (1960), reviewing the literature, states that a number of authors have suggested that schizophrenics with a leptomorphic body build tend to have an earlier age of onset and show a greater degree of withdrawal, apathy, and thought disorder, whereas schizophrenics of euryomorphic body build tend to have a later age of onset, better preservation of personality, and better affective relations with the environment.

There is also some evidence to suggest that physique may have some bearing on the course and prognosis of a schizophrenic illness. Kallman (1953) found that high linearity and low athletic components in physique are associated with a tendency to marked deterioration during the course of the schizophrenic illness, whereas patients with high athletic and low linear components do not show such deterioration; these last are usually paranoid schizophrenics.

Kallman (1953) found that monozygotic twins who both suffer from schizophrenia and are of similar body build tend to have similar types of schizophrenic illness with onset at the same age and similar outcomes. When there are differences in physique, there are also differences in the age of onset as well as in the severity of schizophrenic symptoms. In monozygotic twins discordant with regard to schizophrenia, the twin remaining completely free from schizophrenia had significantly greater physical strength and was higher in body weight from early childhood. [See MENTAL DISORDERS, article on GENETIC ASPECTS.]

Parnell (1958) found that the *Lf* type has the longest hospital stay and also tends to onset before



the age of 25, whereas the *Mf* type stays in the hospital the shortest time and least often succumbs to schizophrenia. [See SCHIZOPHRENIA.]

### Psychosomatic disorders

Investigations relating physique to physical and psychosomatic disorders have generally been handicapped by methodological difficulties, mainly the absence of appropriate control groups and failure to control for age. However, some of the better studies are discussed here.

**Peptic ulcer, asthma, gall bladder disease.** Wretmark (1953), in a carefully planned and adequately controlled study, found that duodenal ulcer patients were more linear in physique than a control group and that the more linear the body build, the worse was the prognosis.

Brouwer (1957) was unable either to replicate the findings of Draper et al. (1944), who claimed that mesomorphic body build is characteristic of ulcer patients, or to support the views of Sheldon (1940) that endomorphic mesomorphy characterizes them. Brouwer found, in agreement with Wretmark, that duodenal ulcer patients are more inclined to leptosomatic body build but do not differ significantly in ectomorphy, as measured by Sheldon's system, from healthy male controls or from groups of male patients with other disorders.

Draper et al. (1944) considered that gall bladder disease is associated with a broad, stocky type of physique. Sheldon (1940) also found that gall bladder patients are mesomorphic-endomorphic. Age was not adequately controlled and Brouwer (1957), using Sheldon's method, found a preponderance of endomorphy in female gallstone patients, though not significantly more than in the comparable healthy female control group.

Rees (1957) and Brouwer (1957), using quite different methods, found that asthma is associated with a variety of somatotypes and patients do not differ significantly in physique from the normal population.

**Cardiovascular disorders.** Rees (1963) reported that military patients at Mill Hill suffering from effort syndrome, cardiac neurosis, and neurocirculatory asthenia were significantly more leptomorphic in body build than normal groups. In a comparison of effort syndrome patients, Rees (1944) found that the majority of leptomorphic patients had suffered from effort intolerance and autonomic dysfunction from an early age and had long-standing personality traits of timidity, anxiety, and neuroticism. Eurymorphic patients, on the other hand, were characterized by short duration of illness, which tended to be precipitated by a great

number of exogenous factors, such as severe environmental stress and severe enemy action. In other words, leptomorphic effort syndrome patients exhibited a predominance of constitutional factors, whereas eurymorphic effort syndrome patients had only recently shown effort intolerance, usually brought on by extrinsic factors such as traumatic experiences and febrile illness.

Gertler, Garn, and Levine (1951) showed that endomorphs have a higher level of blood cholesterol than ectomorphs, both in the normal population and among patients suffering from coronary disease. [See PSYCHOSOMATIC ILLNESS.]

### Delinquency and criminality

Glueck and Glueck (1950), in a study of 500 delinquent boys ranging in age from 11 to 18 years, found that in comparison to the normal control group, the delinquents were more mesomorphic and less ectomorphic.

Epps and Parnell (1952) studied a group of 177 women undergoing Borstal training and compared them with a group of 123 university women similar in age distribution. They found that the delinquents were heavier in body build, more muscular, and had more fat.

These results are in good agreement with the Rees-Eysenck findings of a correlation between extroversion and eurymorphic body build and Eysenck's theory linking extroversion with psychopathy and criminality (Eysenck 1960).

During recent times, the study of the human constitution has advanced significantly as a result of the introduction of new mathematical techniques and improved methods of recording and measuring constitutional attributes. This has led to the discovery of efficient and valid methods of assessing various aspects of physique.

The relationships between physical attributes and personality, health, disease, and abnormality of behavior may prove to have a genetic basis, with the genes responsible for physical attributes influencing the manifestation of the genetic determinant of personality, psychiatric, physical, and psychosomatic disorders.

One of the most urgent needs in this field is for longitudinal studies of size and form combined with genetic, biochemical, and endocrine studies.

W. L. L. REES

[Other relevant material may be found in EUGENICS; GENETICS; MENTAL DISORDERS, articles on BIOLOGICAL ASPECTS, GENETIC ASPECTS, ORGANIC ASPECTS; PHYSICAL ANTHROPOLOGY; RADIATION.]

## BIBLIOGRAPHY

- BENEKE, FRIEDRICH W. 1881 *Konstitution und konstitutionelles Kranksein des Menschen*. Marburg (Germany): Elwert.
- BROUWER, D. 1957 Somatotypes and Psychosomatic Diseases. *Journal of Psychosomatic Research* 2:23-34.
- BURT, CYRIL 1938 The Analysis of Temperament. *British Journal of Medical Psychology* 17:158-188.
- CHILD, IRVIN L.; and SHELDON, WILLIAM H. 1941 The Correlation Between Components of Physique and Scores on Certain Psychological Tests. *Character and Personality* 10:23-34.
- CONNOLLY, CORNELIUS J. 1939 *Physique in Relation to Psychosis*. Catholic University of America, Studies in Psychology and Psychiatry, Vol. 4, No. 5. Baltimore: Williams & Wilkins.
- COFFEN, ALEC J. 1958 Psychosomatic Aspects of Preeclamptic Toxaemia. *Journal of Psychosomatic Research* 2:241-265.
- DAHLBERG, GUNNAR 1926 *Twinn Birth and Twins From a Hereditary Point of View*. Stockholm: Tidens.
- DRAPER, GEORGE; DUPERTUIS, C. W.; and CAUGHEY, J. L. JR. 1944 *Human Constitution in Clinical Medicine*. New York: Harper.
- ENKE, W. 1928 Experimentell psychologische Studien zur Konstitutionsforschung. *Zeitschrift für die gesamte Neurologie und Psychiatrie* 114:770-803.
- EPFS, PHYLLIS; and PARNELL, R. W. 1952 Physique and Temperament of Women Delinquents Compared With Women Undergraduates. *British Journal of Medical Psychology* 25:249-255.
- EYSENCK, HANS J. (editor) (1960) 1961 *Handbook of Abnormal Psychology: An Experimental Approach*. London: Pitman; New York: Basic Books.
- GERTLER, MENARD M.; GARN, STANLEY M.; and LEVINE, SAMUEL A. 1951 Serum Uric Acid in Relation to Age and Physique in Health and in Coronary Heart Disease. *Annals of Internal Medicine* 34:1421-1431.
- GIOVANNI, ACHILLE DE (1891) 1909 *Clinical Commentaries Deduced From the Morphology of the Human Body*. London: Rebman. → First published as *Morfologia del corpo umano*: Studi.
- GLUECK, SHELDON; and GLUECK, ELEANOR 1950 *Unraveling Juvenile Delinquency*. New York: The Commonwealth Fund.
- GREGORY, PAUL W.; and CASTLE, WILLIAM E. 1931 Further Studies on the Embryological Basis of Size Inheritance in the Rabbit. *Journal of Experimental Zoology* 59:199-211.
- HOWELLS, W. W. 1952 A Factorial Study of Constitutional Type. *American Journal of Physical Anthropology* New Series 10:91-118.
- HUMPHREYS, LLOYD G. 1957 Characteristics of Type Concepts With Special Reference to Sheldon's Typology. *Psychological Bulletin* 54:218-228.
- KALLMAN, FRANZ 1953 *Hereditry in Health and Mental Disorder*. New York: Norton.
- KEEHN, J. D. 1953 An Experimental Investigation Into the Validity of the Use of Attitude to Colour and Form in the Assessment of Personality. Ph.D. dissertation, Univ. of London.
- KIBLER, M. 1925 Experimentalpsychologie: Beitrag zur Typenforschung. *Zeitschrift für die gesamte Neurologie und Psychiatrie* 98:524-544.
- KLINEBERG, OTTO; ASCH, SOLOMON E.; and BLOCK, HELEN 1934 An Experimental Study of Constitutional Types. *Genetic Psychological Monographs* 16:145-221.
- KRETSCHMER, ERNST (1921) 1936 *Physique and Character: An Investigation of the Nature of Constitution and the Theory of Temperament*. London: Routledge. → First published in German.
- LINDBERG, BENGT J. 1938 *Experimental Studies of Colour and Non-colour Attitude in School Children and Adults*. Acta psychiatrica et neurologica scandinavica, Supplement No. 16. Copenhagen: Munksgaard.
- LINDEGARD, BENGT 1953 *Variations in Human Body Build: A Somatometric and X-ray Cephalometric Investigation on Scandinavian Adults*. Acta psychiatrica et neurologica scandinavica, Supplement No. 86. Copenhagen: Munksgaard.
- LINDEGARD, BENGT (editor) 1956 *Body-build, Body-function, and Personality: Medical Anthropological Investigations on 320 Swedish 20-year-old Healthy Army Men*. Lund (Sweden): Gleerup.
- LINDEGARD, BENGT; and NYMAN, G. EBERHARD 1956 *Interrelations Between Psychologic, Somatologic and Endocrine Dimensions*. Lund (Sweden): Gleerup.
- MISIAK, HENRYK; and PICKFORD, R. W. 1944 Physique and Perseveration. *Nature* 153:622 only.
- PARNELL, R. W. 1954 Somatotyping by Physical Anthropometry. *American Journal of Physical Anthropology* New Series 12:209-240.
- PARNELL, R. W. 1958 *Behaviour and Physique*. London: Arnold; Baltimore: Williams & Wilkins.
- PILLSBURY, W. B. 1939 Body Form and Introversion-Extroversion. *Journal of Abnormal and Social Psychology* 34:400-401.
- REES, W. L. L. 1943 Physique in Relation to Effort Syndrome, Neurotic and Psychotic Types. M.D. thesis, Univ. of Wales.
- REES, W. L. L. 1944 Physical Constitution, Neurosis and Psychosis. Royal Society of Medicine, *Proceedings* 37:635-638.
- REES, W. L. L. 1950a A Factorial Study of Physical Constitution in Women. *Journal of Mental Science* 96:620-632.
- REES, W. L. L. 1950b Body Build, Personality and Neurosis in Women. *Journal of Mental Science* 96:426-434.
- REES, W. L. L. 1950c Body Size, Personality and Neurosis. *Journal of Mental Science* 96:168-180.
- REES, W. L. L. 1957 Physical Characteristics of the Schizophrenic Patient. Pages 1-14 in Derek Richter (editor), *Schizophrenia: Somatic Aspects*. Oxford: Pergamon.
- REES, W. L. L. (1960) 1961 Constitutional Factors and Abnormal Behaviour. Pages 344-392 in Hans J. Eysenck (editor), *Handbook of Abnormal Psychology: An Experimental Approach*. London: Pitman; New York: Basic Books.
- REES, W. L. L. 1963 An Appraisal of the Concept of Cardiac Neurosis. *Acta psychotherapeutica et psychosomatica* 11:242-252.
- REES, W. L. L.; and EYSENCK, HANS J. 1945 A Factorial Study of Some Morphological and Psychological Aspects of Human Constitution. *Journal of Mental Science* 91:8-21.
- REYNOLDS, EARLE L. 1943 Degree of Kinship and Pattern of Ossification. *American Journal of Physical Anthropology* New Series 1:405-416.
- SANFORD, R. NEVITT et al. 1943 Physique, Personality and Scholarship: A Cooperative Study of School Children. Society for Research in Child Development, *Monographs* 8, no. 34.



- SELTZER, CARL C. 1946 Body Disproportions and Dominant Personality Traits. *Psychosomatic Medicine* 8: 75-97.
- SELTZER, CARL C.; and BROUHA, LUCIEN 1943 The "Masculine" Component and Physical Fitness. *American Journal of Physical Anthropology New Series* 1:95-104.
- SHELDON, WILLIAM H. 1940 *The Varieties of Human Physique: An Introduction to Constitutional Psychology*. New York: Harper.
- SHELDON, WILLIAM H. 1942 *The Varieties of Temperament: A Psychology of Constitutional Differences*. New York: Harper.
- SMITH, H. C.; and BOYARSKY, S. 1943 Relationship Between Physique and Simple Reaction Time. *Character and Personality* 12:46-53.
- TANNER, JAMES M. 1951 Current Advances in the Study of Physique: Photogrammetric Anthropometry and an Androgyny Scale. *Lancet* [1951] no. 1:574-579.
- TANNER, JAMES M. 1955 *Growth at Adolescence*. Springfield, Ill.: Thomas.
- TANNER, JAMES M. 1960 Genetics of Human Growth. Volume 3, pages 43-57 in *Symposia of the Society for the Study of Human Biology, Human Growth*. Edited by James M. Tanner. Oxford: Pergamon.
- WRETMARK, GERDT 1953 *The Peptic Ulcer Individual: A Study in Heredity, Physique, and Personality*. Acta psychiatrica et neurologica scandinavica, Supplement No. 84. Copenhagen: Munksgaard.

## IV

## EXISTENTIAL PSYCHOLOGY

Existentialism is a broad cultural movement showing itself in almost all aspects of contemporary culture: literature, art, philosophy, theology. We can best see the meaning of this development in psychology and psychiatry by first defining the term.

"Existence" comes from the Latin root *ex-sistere*, which means literally "to stand out," "to emerge." Every being, whether a tree or dog or man, "stands out" from being in general to the extent that it has individual existence. But the human being is characterized by his capacity to be *aware* of the fact that he stands out; that is to say, I can be conscious of the facts that I am this particular man (middlewestern American) who stands at this particular point in history (mid-twentieth century), who exists in a given physical and social world (New York City), who has the intention at this moment of writing this article, and so on. The term our German colleagues use for the human individual is *Dasein*, composed of *Sein*, "being," plus *da*, "there." Man's capacity to become aware of his existence—which distinguishes him as a human being, separate from the rest of nature—implies that he can *respond* to the facts of his own existence and thus exercise some responsibility for it.

**Existence versus essence.** In Western thought there are traditionally two philosophic concepts—"existence" and "essence"—customarily used in contrast to each other. "Essences" refer to discrete entities which can be separated out and formulated, such as Plato's "ideas" or Hegel's panrationalistic concepts. In science, essences have come to refer to attributes such as weight, density, color, or in the study of the human being to neurological patterns, reflexes, drives, and so on. By and large, Western thought since the Renaissance has been concerned with essences; traditional science, assuming an essentialist metaphysics, has sought to discover these substances and discrete relationships. The existential psychologists and psychiatrists hold, in contrast, that if we try to understand man as a bundle of discrete drives or a composite of reflex patterns, we may end up with brilliant generalizations but we have lost *the man* to whom these things happen, the human being we set out to understand in the first place. They are obviously not against the study of dynamisms, neurological and chemical relationships, and patterns of behavior as such. But they hold that these cannot be understood in any given individual—such as a patient in therapy—except in the context of the overarching fact that here is a person who *exists*, who is; and how this person relates to his existence at this given moment is essential to the meaning of a given pattern of behavior. The existential approach is thus dynamic in the fundamental sense that it takes the person always as emerging, always in process of becoming. The term "being," indeed, is not a static noun but a verb form.

Thus the existentialist view of science is close to that of the modern physicists Niels Bohr and Werner K. Heisenberg, who insist that the Copernican separation of subjective man from objective nature is no longer tenable and that the ideal of a natural science which is completely objective is an illusion. Also it is to be noted that the contemporary existential emphasis on being is, in one sense, a rediscovery in the West of emphases which Eastern thought has held for centuries but which have been lost in Western overrationalistic and objectivistic concerns since the Renaissance.

## History

Existential psychology has many illustrious progenitors in Western history, of whom we shall mention only three. Socrates, whom Kierkegaard later took as his model, is the prototype of the existentialist in his method of dialogue and his belief that truth is discovered through rigorous, inward

self-examination. Augustine is prominent in this heritage in his concern with will and with psychological time. Pascal, a distinguished scientist who saw clearly the limits of natural science for understanding man, emphasized the human being's *affective* grasp of truth: "The heart has reasons which the reason knows not of."

Søren Kierkegaard. Modern existential psychology took its specific form a hundred years ago in Kierkegaard's passionate attack on Hegel, in whose philosophy, as in other aspects of nineteenth-century culture, essentialism came to its peak. Kierkegaard proclaimed that Hegel's identification of reality with abstract concepts was trickery and that it seduced men away from their central concern, which was their own understanding and their own decisive actions. Kierkegaard and the existentialists following him (Schopenhauer, Marx, Dilthey, and particularly Nietzsche) fought against the rationalists and idealists who would see man only as subject. But even more strongly they attacked the opposite tendency which was emerging in the nineteenth century to become dominant in the twentieth—namely, the vast pressures to treat man as an object to be calculated, manipulated, and controlled. They saw this trend developing in the application of science to man, and they feared this would sap man's capacity for decision and individual responsibility. The trend was also present in the growing industrialism and capitalism, where it was attacked by Marx in his early existential statements asserting that the money economy would dehumanize man. The distinctive quality of *revolt* which characterizes existentialists arises from their common protest against these tendencies toward the objectification of man.

William James. The American psychologist William James was a participant in the existential development; he came back from Europe in the latter part of the nineteenth century protesting against Hegelian rationalism just as Kierkegaard had protested half a century earlier. James's existential elements include, first, his passionate emphasis on the *immediacy of experience*. Second, he emphasized *will and decision*; he held that commitment is a prerequisite to the discovery of truth. Third, like Kierkegaard, he distrusted thought separated from action. James's epistemology has striking similarities to that of Nietzsche in the first part of the *Will to Power*, where Nietzsche held that truth is the way a biological group actualizes itself. Finally, James's devotion to social and humanistic values as the overarching context, the "enveloping wider order," in which natural science

plays only one part, makes him similar to the existentialists; he was thus able to bring art and religion into his psychological thought without sacrificing his scientific integrity. James was also, of course, in the natural science tradition, and it is unfortunate that, although his natural science orientation was followed, these broader emphases were ignored or treated with contempt until recently. [See the biography of JAMES.]

Sigmund Freud. Freudian psychoanalysis is also, on one side, an expression of what we have broadly designated as the existential development. Freud stands in the line of those explorers of human nature of the nineteenth century—including Kierkegaard, Nietzsche, Schopenhauer—who rediscovered the significance of the irrational, dynamic, "unconscious" elements in personality which had been repressed by the essentialistic presuppositions of Western thinking since the Renaissance. Freud sought to bring the irrational springs of human behavior back into the accepted area of scientific investigation. At the same time, Freud saw himself as an enthusiastic champion of natural science, his goal being to make the irrational elements explicable in terms of the mechanistic Newtonian model in nineteenth-century physics and biology. The existentialists radically disagree with this model of man. In their famous discussions, Binswanger (1956) endeavored to show his friend Freud that the latter's determinism was a metaphysical assumption and needed to be analyzed like any other assumption. But while Freud held resolutely to his determinism in his theoretical model of the mind, in his therapeutic practice he inconsistently remained to a great extent existential. [See the biography of FREUD.]

### Contemporary contributions

The existential approach sprang up among men of diverse views in various parts of Europe beginning in the early part of the twentieth century. Eugene Minkowski in Paris, Erwin Straus in Germany and later in the United States, and V. E. von Gebattel in Germany represent chiefly the first, or phenomenological, phase of this development. The psychiatrist Karl Jaspers, who was later to become a leading German philosopher, wrote a textbook on psychopathology in which he presented a phenomenological analysis of the different kinds of pathology. The book went through many editions and widely influenced psychiatry and research on the Continent; it is now available in English (Jaspers 1913). The second phase, which is more specifically existential, is represented by Ludwig



Binswanger, A. Storch, Medard Boss, Gustav Bally, and Roland Kuhn in Switzerland, F. J. Buytendijk and J. H. Van den Berg in the Netherlands, and others.

The fact that this development emerged spontaneously—some of these existential thinkers often scarcely knowing about the remarkably similar work of their colleagues—and the fact that it owes its creation to diverse psychiatrists and psychologists rather than to any one leader, argue that it is a response to a widespread need in our time. The variety of schools can be seen when we note that Minkowski and Straus, for example, were psychiatrists and neurologists, not psychoanalysts; von Gebattel, Boss, and Bally were Freudian analysts; Binswanger, although remaining in Switzerland, became a member of the Vienna Psychoanalytic Society at Freud's recommendation when the Zurich group split off from the international group; other existential therapists had also been under the influence of Carl G. Jung.

All of these men shared grave doubts about the theory of man assumed in contemporary psychiatry and psychology, and specifically in psychoanalysis. They had become disquieted over the fact that although they were effecting cures by the techniques they had learned, they could not, as long as they confined themselves to the generally accepted assumptions, arrive at any clear understanding of why or why not these cures occurred. They were unwilling to postulate unverifiable agents such as "libido," or "censor," and they asked trenchant questions about what it really means when one says a patient "projects" or "transfers" something to someone else. These explanatory devices, they argued, as often as not actually cover up rather than reveal precisely the empirical data in the patient we wish to understand. They shared the belief that the serious difficulties in the accepted theories of man, including psychoanalysis, would not only gravely limit research in the long run but would also seriously hamper the effectiveness and development of therapeutic techniques.

These psychologists and psychiatrists were also concerned with events *prior* to the observed patterns of behavior; they asked questions about how particular kinds of behavior are made possible in the patient's experience. Thus their approach was chiefly concerned not with types of pathology but with norms and structures of being human; the patient's pathology is a distortion of these. They used psychological terms with an ontological meaning. Kierkegaard held that anxiety is ontological in that it is a constitutive element for the self; a person never actualizes a potentiality without go-

ing through what Kierkegaard called the "intermediate determinant" of anxiety (1844). Similarly, Tillich described anxiety as the state of the struggle of "being with non-being" (1952). Nietzsche's "will" is ontological in that it is necessary for the individual's self-actualization. Tillich also made "courage" ontological in his trenchant phrase "the courage to be."

In their therapy, the existential psychologists and psychiatrists also asked such prior questions. Transference, they agreed, is one of Freud's greatest discoveries, but it can be understood only on the basis of the prior norm of *encounter*, of which transference is a distortion. As the existentialists describe it, encounter includes not only sexual elements (*eros*) but such normal elements as mutual esteem and concern between persons (*agape*) and realistic friendliness (*philia*). Likewise, Freud's concepts of resistance and repression refer to observable behavior, but they can be understood only as distortions of the patient's freedom, that is, of his capacity to actualize some awareness or experience, a freedom which assumedly arouses too much anxiety for the patient to face at the moment.

Admittedly, these ontological questions sound alien to most American psychological thinking, accustomed as it is to "essentialist" behavioristic presuppositions. But research is now going on, to which we shall refer below, which not only may make a bridge to various forms of psychology, but which can bring new and significant empirical data to bear upon our psychology.

### Central theoretical orientations

Prior to presenting concepts, it is valuable to see the place of existential psychology in relation to psychology as a whole. Van Kaam (1966) distinguishes between "differential psychologies," whose concern is the observation and explanation of isolated behavioral processes and functions, and "comprehensive psychologies," whose aim is the full understanding of human behavior as a whole and in terms of its uniquely human characteristics. Most psychologies have been differential psychologies and have not been integrated into a total image of man. Existential psychology is a comprehensive psychology whose aim is an integration of the observations of differential psychologies into an explanatory theory about human behavior in its lived intentional entirety.

Existential psychologists and psychiatrists share several basic theoretical orientations. First, they deny the subject-object dichotomy that has been dominant in Western thought since Descartes. They emphasize, in contrast, that man exists as

subject and object at the same time. May (1966a) holds that a dialectical relationship between subjective and objective factors occurs in every human action and that in this relationship lie the sources of human freedom and creativity. Every act of consciousness is consciousness of something; that is, it has its objective as well as its subjective pole. Consciousness is an active process; the man who is doing the experiencing and the objective world which is the object of consciousness can never be separated.

Second, these psychologists and psychiatrists share a method of scientific investigation—phenomenology—which undercuts the usual objectification in traditional science, employing an unbiased approach.

As Straus points out in his *Phenomenological Psychology*, science takes for granted precisely what a true psychology must focus upon, namely, our capacities to see, to observe, and to describe. The task of the phenomenologist is less to examine the things observed and more to reflect upon the situation of the observer. The structure and constitution of everyday life (the *Lebenswelt*) becomes the object of examination because it is this that makes possible the very acts of seeing, observing, and describing, and that gives them their uniquely human qualities. It can be questioned whether phenomenology in psychology and psychiatry is "absolutely" unbiased. For the mere act of listening to a patient in therapy or to a subject in an experiment presupposes some concepts and symbols by which we listen. Nevertheless, the experience of many psychotherapists is that the phenomenological method represents a radical "breakthrough" as a method for understanding the human being. It raises "participant observation," to use Sullivan's phrase, to a broader dimension and gives it a structural base as an epistemology.

A third concept, "being-in-the-world," is hyphenated to indicate again that the being and its world cannot be separated. On the human level world is defined as the meaningful structure of events in which the human being experiences himself; self implies world, and world self, and one is not possible without the other. Furthermore, three modes of world are distinguished. *Umwelt*, literally the "around world," is what is customarily called environment, the "natural world." *Mitwelt*, literally the "with-world," is the world of one's relationships with his fellow man, a description on a broader structural level of the world of "interpersonal relations." Buber has made a significant contribution to understanding *Mitwelt* through his analysis of "I-thou" relationships (1936). *Eigenwelt*, literally

the "own world," has to do with the self's relation to itself.

A fourth theory in existential psychology and psychiatry has to do with the time dimension. Heidegger's great contribution, writes Straus (1936), was to "vindicate time for existence," and he adds that time is to psychology what mathematics is to physics. Existential psychotherapists, following Heidegger here (1927), place particular emphasis on how the patient relates to the future time dimension: they do not rule out concern with the past but see it as having meaning in terms of the present and the future. The depressed patient, as Minkowski has indicated (see May et al. 1958, p. 127), is characterized by being unable to live in the future, unable to hope, or to project himself into the morrow. Minkowski reverses the usual explanation that the patient cannot hope because he is depressed and proposes that the more useful approach explains that the patient is depressed because he has lost the capacity to live in this future dimension of time. May (1966a) points out that in therapy the acceptance and working through of the subject-object dilemma gives the patient a new sense of freedom with respect to time and the capacity to use time in the service of self-chosen goals. [See TIME, article on PSYCHOLOGICAL ASPECTS.]

A fifth common theory, already alluded to, emphasizes choice, responsibility, and freedom. As Laing puts it, "... the task in psychotherapy [is] to make ... an appeal to the freedom of the patient" (1960, p. 64). Freedom is real, not illusory, and it derives from the self-constitutive powers of the *Dasein*. I am the *object*, for example, of conditioning, of bodily factors, of limitations of time and space, to say nothing of limitations of understanding and perceptions and all the outside forces determining me. But insofar as I am conscious of these facts I am *subject*, and I can, and to some extent must, choose the attitude I will take toward them. A person exercises choice in structuring his modes of being and his world. He bears the responsibility for his being-in-the-world even as it manifests itself in his dreams; in other words, one is responsible even where least free. Sartre (1943) has made the famous statement "existence precedes essence," which in effect implies that the human being has unlimited freedom to transform himself. Such an extreme existentialism, however, is not held by adherents of the other existential theories, who emphasize that both essentialist and existentialist elements are necessary for any comprehensive psychology. Existential psychotherapy by no means ignores the infinite number of deterministic elements in the patient's life and



problems, but it believes that helping the patient explore and become aware of these deterministic elements enlarges the sphere of the patient's freedom to choose his relationship to these determining facts. The acceptance of limitations is thus not a negative process but itself gives the basis for creative self-actualization.

Maurice Friedman has compiled a comprehensive and critical anthology of selections from the whole range of existentialist writings. Organizing the material around the central issues rather than around authors, Friedman is able to highlight the differences between disparate philosophies of existentialism, but more importantly, to demonstrate the common emphasis on the dynamics, the concrete, personal involvement and commitment. Furthermore, he makes clear the area of intersubjectivity, which is most crucial for understanding similarities and differences among contemporary existentialists. "A distinction must be made between such thinkers as Heidegger, Sartre, and Tillich, for whom the intersubjective tends to remain a dimension of the self, and such thinkers as Buber, Marcel, Jaspers, and Camus, who in one way or another see the relations *between* man and man as central to human existence" (Friedman 1964, p. 173).

### Current research

We shall mention a variety of research studies, the first having to do with precisely the problem we have just been discussing. It is represented by a recent study by Vera Gatch (1963) on the effect of philosophical commitments to psychic determinism or freedom on the behavior of the psychotherapist. Gatch matched a group of ten existentialist therapists against ten orthodox Freudian therapists who were committed to the theory of determinism. One highly significant difference she found between these two groups was that the existential therapists referred to issues of choice, decision, and responsibility with their patients much more frequently than did the Freudian therapists.

Another type of research is Erwin Straus's classical work on sensory experience. Straus shares a major concern of the phenomenological-existential psychiatrists and psychologists, which is the extent to which traditional theoretical commitments implicitly structure experimental investigations and consequently predetermine the results. "Experiments alone do not guarantee pure empiricism," he writes ([1936] 1963, p. 4) and goes on to point out how contemporary empirical, "objective" psychology is shot through with limiting and distorting metaphysical biases. In an extensive and detailed analysis of Pavlov's work and its Cartesian

underpinnings, Straus demonstrates that Pavlov ignored the basic difference between the relationship of an organism to stimuli and of an experiencing being to the world of objects (p. 21). In reducing sensing to epiphenomena of physiological processes occurring within an organism, violence is done to the observable facts. The essence of sensory experience is its intentionality. Sensory impressions imply questioning, seeking, and expecting orientations to the world (p. 101).

Straus forcefully demonstrates the inadequacy of conditioning theory, as well as its theoretical derivatives, in explaining even the behavior of Pavlov's dogs, let alone the behavior of humans. His critique is based on a careful phenomenological study of the descriptions of the dogs' behavior given by Pavlov himself. Straus shows that Pavlov was highly selective, since he chose to deal with only those aspects of the total experimental situation that fit his philosophical presuppositions. Straus's concern, however, is not primarily to criticize Pavlov. His major aim is to reclaim sensory experience from the reductionism of physics and physiology and to restore it to its central position in a psychology of human experience.

In a similar vein, Joseph Lyons (1963) has set himself the task of re-evaluating fundamental concepts of contemporary clinical psychology. He claims that psychology has been dominated by the intention to predict and control behavior rather than by an interest in understanding the reality of the human person. A scientific psychology of people can only be established on a phenomenological basis and requires two conditions: (1) that the perceived be a person, not an "impersonal other," and (2) that the perceiver act as a full human being. Psychology then becomes "the science of persons as they are apprehended by other persons." Lyons analyzes a host of concrete, traditional clinical problems in the context of this perspective, emphasizing, as does Straus, the central role of intentionality. For example, he argues that even the stimulus in the typical stimulus-response experimental situation partakes of the subject's intentional state. The stimulus is given special meaning by virtue of the subject's readiness for it, and a full description of such a stimulus must take that into account.

Another type of research going on at several universities is the study of subjects' responses, on the Rorschach or similar tests, not with the question of what the response itself means, but with the question of what it means that the subject is able to give such and such a response. The question, then, is not simply *how* does the subject structure his world, but what is occurring that enables him to structure his world this way or that, and

what decision and selective processes go on in this structuring. Such studies are examining the *intentionality*—in the phenomenological-existential sense—that goes into responses and the *potentiality* that makes the responses possible. Caligor and May (1967) made a study of the intentionality shown in the symbols in one patient's dreams through an entire psychoanalysis. They propose that a clearer and more trustworthy prediction of what is occurring and going to occur in the patient's therapy could be made from noting how the patient forms and reforms his symbols at various stages in the therapy than from the patient's conscious verbal communications with the therapist.

Ronald Laing (e.g., Laing & Esterson 1965), a leading exponent of the applications of the phenomenological approach to psychoanalysis, has been studying the family interaction processes of schizophrenic patients. Combining the interpersonal and phenomenological frameworks, Laing maintains that schizophrenia becomes intelligible as the result of ambiguous and contradictory environmental processes. It is a protective device for living in an otherwise hopelessly confusing environment—a strategy for survival. Laing avoids the concepts of "disease" and "pathology" and demonstrates that what happens in a group will be intelligible if one can retrace the steps from what is going on to who is doing what. His work suggests modifications of traditional views of such phenomena as regression, acting out, catatonia, and transference.

The focus of study for Leslie Farber (1966) has been a psychology of will, a subject notably absent in psychological theory and research. Farber observes that although we recognize the presence of many causes for any behavior, we attribute the motivating power to a single aspect of the causal nexus, such as libido, sexuality, or anxiety. However, the only prime mover that is held accountable, answerable, is will, man's power of volition. When a personality theory does not deal with an explicit, accountable will, it smuggles it in under another name and makes it an irresponsible, that is, unaccountable, prime mover. The focus then shifts to motives as the prime causes of behavior. As a consequence, the phenomenological understanding of both will and motives suffers. Farber points out that will and motive are superficially similar in that both may be involved in provoking action. But a motive cannot be responsible for an action of will. For example, acting disparagingly may be motivated by envy. But since one could also choose not to disparage, even though envious, it is will that is responsible for the action, not envy. Farber concretely applies his concept of the privileges and

limitations of the will to a range of clinical issues such as anxiety, schizophrenia, and sex.

### Existential analysis ("Daseinsanalyse")

Existential analysis, in contradistinction to existential psychotherapy, is a type of phenomenological, empirical investigation of specific and actual forms and patterns of existence. In other words, given the essential structure of human nature described by Heidegger (1927) as "being-in-the-world," the existential analysts, among whom are the Swiss psychiatrists Binswanger and Boss, direct their attention to the particular, concrete modes of being-in-the-world of a particular person. They attempt to *understand* (in contrast to *explaining*) the subjectively experienced complex of meanings that constitute a given person's world designs.

**Understanding versus explanation.** It is in the contrast between understanding and explanation that the difference between existential analysis and psychoanalytic theory may be explicated. Understanding involves comprehending the object on its terms,

to see in it structures that emerge from *its* side, and not from ours. . . . In explanation, the phenomena as they appear are transformed in the sense that they are either subsumed under laws that relate them to other, different phenomena, or they are broken down into parts that somehow are taken as more real than the configuration of those parts that are taken to make up the phenomenon in question. (Binswanger [1922–1957] 1963, pp. 38–39)

In this sense psychoanalytic theory is an explanatory system, and Freud's conception of man as *Homo naturalis*, that is, as essentially a biophysical organism, places psychoanalytic theory within the domain of natural science. As such, the "psychic apparatus" is a derivative of biological instincts, and the experienced world is *explained* by transforming it into its biological components. Binswanger pays great tribute to Freud's genius in developing the first thoroughgoing natural science view of man. His criticism of Freud is not that his theoretical structure is erroneous but that it is limited. Psychoanalytic theory deals with but *one* mode of being-in-the-world, namely, what Heidegger calls "thrownness." The mode of "thrownness" is that which is structured or determined by one's biological and constitutional make-up (the *Umwelt*) and conditioned by one's past experience. It is the mode of being-in-the-world without freedom.

**Psychopathology.** Although existential analysis examines the person and his world without regard to criteria of health or pathology, the very notion of the essential structure of *Dasein* generates a concept of psychopathology. The essence of *Dasein*



is openness to multiple modes of being-in-the-world and readiness to experience meaning in these multiple modes. Any modification of the essential structure that results in "constriction," "narrowing," or "flattening" of the world, so that experience can be organized in only one, or a few, categories, involves a loss of existential freedom and is therefore pathological. Thus far, the only world designs that have been subjected to existential analyses are severely constricted ones. The published existential analytic case studies of both Binswanger (1956; 1962) and Boss (1957; 1962) are all of psychotic or severely neurotic patients. It is hoped that analyses of non-constricted modes of being-in-the-world may shed light on the structure of creative world designs.

**Creative phenomena.** The existential analysts have always showed special interest in creative and spiritual phenomena. They believe that these kinds of phenomena cannot be explained in terms of antecedent or biological components, for the essence of the phenomenon is precisely the emergence of the new. The essence of the creative man is his being-in-the-world in freedom. The existential analysts similarly emphasize that in therapeutic practice a total commitment to psychic determinism rules out seriously entertaining the possibility of freedom of choice and leaves one only able to foster "the illusion of freedom." Whether or not existential analysis will shed light on the problems of creativity remains to be seen, but at least its methods open up the possibility. We should add that some analysts in the Freudian tradition, and some others who go under the ambiguous rubric "ego psychologists," do expand the presuppositions of psychoanalytic theory, implicitly and explicitly, and have made contributions which have something in common with the existential approach.

### Psychotherapy

Binswanger and Boss make a sharp distinction between psychoanalytic theory and psychoanalytic therapy. They have much to say about the limitations of the theory, but they value and use the therapy within a new context. Binswanger actually concerns himself less with psychotherapy, being interested chiefly in the "philosophical anthropology" on which therapy is based. Aside from a few descriptive references to the task of therapy as a process of assisting man in his ability to structure his world and a statement about the therapist's orientation toward encounter with, and understanding of, the whole person, Binswanger accepts Freudian therapy as a potent and effective clinical instrument.

Boss, on the other hand, deals extensively and explicitly with psychotherapeutic practice. Whereas

Binswanger's case histories are exclusively analyses of modes of being-in-the-world, Boss's case studies are cast in the framework of the therapeutic process. Furthermore, Boss undertakes a detailed examination of the procedures of Freudian therapy and concludes that Freud the therapist is radically different from Freud the mechanistic theorist.

Although Boss's case studies are instructive in casting traditional psychodynamic concepts into the broader, more meaningful phenomenological framework of existential analysis, there seems to be an oversimplification in both Boss's image of man and his view of psychotherapy. For example, Boss claims that all anxiety can be removed by love and trust, whereas one of the major contributions of existential thought is its recognition of anxiety as an ontological condition of existence.

What does the designation "existential psychotherapy" mean? Certainly it does not refer to a self-contained depth therapy that is counter to Freudian psychoanalytic psychotherapy, as is the therapy of Jung or Reich. Nor does it refer to a body of special techniques or procedures, although it has made and will presumably make significant contributions to technique. It is more accurate, indeed, to speak of the *existential attitude toward, or context of, psychotherapy*, since existentialists have thus far focused their attention on the presuppositions underlying all psychotherapies. There are, however, a number of attributes that are descriptive of the existential attitude. Existential psychiatrists and psychologists believe that this attitudinal framework—which is the application to clinical practice of theories mentioned earlier—fosters psychotherapeutic functioning which is more directly relevant to the problems of human beings in our contemporary age and is more consistently efficacious.

The first attribute may be described as observation in the phenomenological mode. That is, the therapist must learn to perceive the patient in terms of his (the patient's) phenomenal reality and not in terms of the therapist's theoretical models. The therapist can never know the patient's experiential world in the same sense that he can know his own. The critical problem is that of participation in the patient's world, whether one calls the means by which this is done empathy, intuition, partial identification, presence, or encounter. This is obviously such an essential process in any psychotherapy, indeed in all kinds of human interaction, that it is surprising it has not been subjected to research investigation. Hopefully, a future area of psychotherapy research will be the study of the therapist's inferential processes.

A second attribute is the experiential mode of

knowing. "A psychotherapy on existential-analytic bases," writes Binswanger, "thus proceeds not merely by showing the patient where, when, and to what extent he has failed to realize the fullness of his humanity, but it tries to make him experience this as radically as possible . . ." (1962, p. 20). The experiential mode applies to the therapist as well. An aspect of his "presence" in the "encounter" with the patient is his experiencing the patient's communications on many different levels of his own being.

Third, the existential attitude emphasizes a primary, but not exclusive, focus on the "here and now." It is destructive of the meaningfulness of existence to interpret essential potentialities as solely genetic developmental processes. A person's whole history is built into all present relationships and is one of the constitutive determinants of them. But to say that the past is the sufficient cause of the present and to examine the past in therapy in this context is an instance of the kind of reductionism the whole existential movement adamantly opposes. At the same time, we caution against the omission of the genetic dimension in the phenomenological approach. Such an omission points to one of the shortcomings of existential analysis—namely, its lack of treatment of psychological development. Some phenomenological purists claim that the developmental processes can never be the subject of phenomenological investigation, but in our opinion a phenomenology of human development makes sense and has in fact been demonstrated in the illuminating studies by Schachtel (1959). In any case, the therapeutic consequences of verbalizations about the patient's past history is a question which suggests itself as a fruitful area for research. In practice, many therapeutic schools seem to be moving toward emphasis on the patient's present situation. It is interesting to note that in her study comparing Freudian and "existential" analysts Gatch found that *neither* group made many references to the patients' previous experiences, nor did they make many causal interpretations.

Fourth, the existential approach keeps intentionality always at the heart of the therapeutic process. May (1966a) points out that intentionality is shown, in one aspect, in the movements and gestures of the body; thus the body and "instincts" come directly into existential therapy, not as a set of discrete drives but as the totality of movement of the organism. Intentionality is seen as the source of will and decision and also as the area in which human wishing, including the Freudian wish, operates. This central emphasis on intentionality has opened up some areas for constructive therapy

which were previously thought to be resistive to all psychotherapy, such as drug addiction.

Bugental (1965) has attempted to elucidate the meanings of the concepts of existentialism within the framework of psychotherapeutic practice. He sees the therapeutic task as the pursuit of authenticity, and this task has two phases. The analytic phase is reparative and is essentially an attack upon the distortions of awareness and being that cripple a person's existence. The second phase, the ontologic, "is concerned with the realization of the potential for liberated being" (1965, p. 183).

We conclude with a comment that arises out of the fact that existentialists, because of their commitment to the "inviolability of human experience" and to the phenomenological method of investigation, have tended to eschew systematization and scientific theory construction (Ford & Urban 1963, p. 480). But both approaches are necessary to the advancement of knowledge. Neither scientific conceptualization nor experimentation as such need violate meaningful experience. The best safeguard against their doing so is a thorough clarification of their implicit presuppositions. Existential analysis has insisted on such clarification. It may also construct thereon a broader humanistic scientific theory, as Binswanger and Straus have already begun to do.

ROLLO MAY AND SABERT BASESCU

[See also PHENOMENOLOGY. Other relevant material may be found in ANXIETY; MENTAL DISORDERS, TREATMENT OF, article on CLIENT-CENTERED COUNSELING; MOTIVATION; PERSONALITY; CONTEMPORARY VIEWPOINTS; PSYCHOANALYSIS; RELIGION; and in the biographies of BUBER and HUSSERL.]

#### BIBLIOGRAPHY

- BASESCU, SABERT 1962 Human Nature and Psychotherapy: An Existential View. *Review of Existential Psychology and Psychiatry* 2:149-157.
- BINSWANGER, LUDWIG (1922-1957) 1963 *Being-in-the-World: Selected Papers*. Translated by Jacob Needleman. New York: Basic Books. → First published in German. Contains a critical introduction to Binswanger's existential psychoanalysis.
- BINSWANGER, LUDWIG (1956) 1957 *Sigmund Freud: Reminiscences of a Friendship*. New York: Grune. → First published as *Erinnerungen an Sigmund Freud*.
- BINSWANGER, LUDWIG 1962 Existential Analysis and Psychotherapy. Pages 17-23 in Hendrik M. Ruitenbeek (editor), *Psychoanalysis and Existential Philosophy*. New York: Dutton.
- BOSS, MEDARD (1957) 1963 *Psychoanalysis and Daseinsanalysis*. New York: Basic Books. → First published in German.
- BOSS, MEDARD 1962 Anxiety, Guilt and Psychotherapeutic Liberation. *Review of Existential Psychology and Psychiatry* 2:173-195.



- BUBER, MARTIN (1936) 1958 *I and Thou*. 2d ed. New York: Scribner. → First published in German.
- BUBER, MARTIN 1957a *Distance and Relation*. *Psychiatry* 20:97-104.
- BUBER, MARTIN 1957b *Guilt and Guilt Feelings*. *Psychiatry* 20:114-129.
- BUGENTAL, J. F. C. 1965 *The Search for Authenticity*. New York: Holt.
- CALIGOR, LEOPOLD; and MAY, ROLLO 1967 *Dreams and Personality Change*. New York: Basic Books.
- FARBER, LESLIE H. 1966 *The Ways of the Will: Essays Toward a Psychology and Psychopathology of Will*. New York: Basic Books.
- FORD, DONALD H.; and URBAN, HUGH B. 1963 *Systems of Psychotherapy: A Comparative Study*. New York: Wiley.
- FRIEDMAN, MAURICE 1964 *The Worlds of Existentialism*. New York: Random House.
- GATCH, VERA 1963 *The Effect of a Philosophic Commitment to Psychic Determinism on the Behavior of the Psychotherapist*. Ph.D. dissertation, Univ. of Oklahoma.
- HEIDEGGER, MARTIN (1927) 1962 *Being and Time*. New York: Harper. → First published as *Sein und Zeit*.
- JASPERS, KARL (1913) 1963 *General Psychopathology*. Univ. of Chicago Press. → First published in German.
- KIERKEGAARD, SØREN A. (1844) 1957 *The Concept of Dread*. 2d ed. Princeton Univ. Press. → First published as *Begrebet angst*.
- LAING, RONALD D. 1960 *The Divided Self: A Study of Sanity and Madness*. Chicago: Quadrangle Books; London: Tavistock.
- LAING, RONALD D.; and ESTERSON, A. 1965 *Sanity, Madness and the Family*. Volume 1: *Families of Schizophrenics*. New York: Basic Books.
- LYONS, JOSEPH 1963 *Psychology and the Measure of Man*. New York: Free Press.
- MAY, ROLLO 1950 *The Meaning of Anxiety*. New York: Ronald.
- MAY, ROLLO (editor) 1961 *Existential Psychology*. New York: Random House.
- MAY, ROLLO 1966a *Psychology and the Human Dilemma*. Princeton, N.J.: Van Nostrand.
- MAY, ROLLO 1966b *Love and Will*. New York: Norton.
- MAY, ROLLO; ANGEL, E.; and ELLENBERGER, H. F. (editors) 1958 *Existence: A New Dimension in Psychiatry and Psychology*. New York: Basic Books.
- SARTRE, JEAN-PAUL (1943) 1956 *Being and Nothingness: An Essay on Phenomenological Ontology*. New York: Philosophical Library. → First published as *L'être et le néant: Essai d'ontologie phénoménologique*.
- SCHACHTEL, ERNEST G. 1959 *Metamorphosis: On the Development of Affect, Perception, Attention, and Memory*. New York: Basic Books.
- STRAUS, ERWIN W. (1936) 1963 *The Primary World of Senses: A Vindication of Sensory Experience*. New York: Free Press. → First published as *Vom Sinn der Sinne*.
- STRAUS, ERWIN W. *Phenomenological Psychology: The Selected Papers of Erwin W. Straus*. New York: Basic Books, 1966.
- TILICH, PAUL 1944 *Existential Philosophy*. *Journal of the History of Ideas* 5:44-70.
- TILICH, PAUL 1952 *The Courage to Be*. New Haven: Yale Univ. Press. → A paperback edition was published in 1959.
- VAN KAAM, ADRIAN L. 1966 *Existential Foundations of Psychology*. Pittsburgh: Duquesne Univ. Press.

## V

## APPLIED PSYCHOLOGY

Although formerly used in a narrower sense to include only business and industrial psychology, today the term *applied psychology* designates the professional practice (as distinguished from the basic science) of psychology in all areas. It covers applied research and service functions in the fields of personnel, engineering, consumer, clinical, and counseling psychology. In addition, some applied psychologists work with members of other professions as consultants or in the training of other professional personnel. Before World War II, psychology was primarily an academic discipline. Today over half of America's psychologists work outside of universities, in such settings as government agencies, industrial organizations, school systems, hospitals, clinics, and counseling centers.

## The science and profession of psychology

Applied psychology does not differ in any fundamental way from the rest of psychology. In terms of orientation and training, every applied psychologist is a psychologist first and an applied specialist only secondarily. All have taken essentially the same types of courses; specialization occurs chiefly at the level of practicum training and on-the-job experience. The clinical, counseling, or personnel psychologist, for example, has received training in experimental psychology, and the experimental psychologist utilizes in his research concepts or hypotheses that may have originated in the clinic or factory.

The unique contribution of the applied psychologist, as distinguished from that of other specialists, stems largely from his *research approach* to problems of human behavior. When the psychologist goes to work in an applied setting, he brings with him not a set of rules or specific facts but a *method* for attacking problems. It is essentially the scientific method, common to all sciences. This method consists in the empirical observation and recording of facts under systematically controlled conditions. By controlling conditions through appropriate experimental designs, the psychologist can identify cause-and-effect relations and can isolate the influence of individual variables. The scientific method also requires replication of observations and a determination of the margin of error inherent in its procedures.

Since applied psychology follows the same fundamental procedures as basic psychological research, how can the two be differentiated? When the applied psychologist is engaged in research—as is true much of the time for many applied psy-

chologists—the distinction hinges on the difference between basic and applied research. This difference is one of degree. Although examples can be found that fall clearly into the basic or the applied category, some research can be classified in either one. One determining factor is the reason for choosing a problem: to help in theory construction (basic) or in making administrative decisions (applied). For instance, is the investigation concerned principally with the nature of learning or with the most effective method for teaching children to read?

Another difference pertains to the specificity or generality of the findings. The results of basic research can be generalized more widely than those of applied research. The latter typically yields results pertaining only to narrowly limited contexts. Similarly, applied research permits less analysis of causal relations, since situations are likely to be compared in their totality rather than being broken down into more elementary components. For example, applied research may demonstrate that one of two training procedures—in all their operational complexities—yields better results than the other in a particular situation. But it may not be possible under these circumstances to specify *why* the one method was superior.

Some applied psychologists spend most of their time in service rather than in research functions. How does such a psychologist use scientific method when engaged, for instance, in psychotherapy with a neurotic patient, in testing a child with a reading disability, in interviewing an applicant for an executive position, or in planning a consumer survey for an automobile manufacturer? He does so in at least two ways. First, he evaluates and chooses his procedures, not on the basis of subjective opinion, but in the light of available published research conducted in comparable situations. Being unable under the circumstances to carry out his own empirical evaluations, he is nevertheless guided by whatever empirical data are available regarding tests, questionnaire techniques, varieties of psychotherapy, or any other procedure he plans to use.

In a still more general way the psychologist's scientific training influences his approach to every problem he encounters in his service functions. Through such training he has formed habits of accurate and objective observation and recording of facts, and he has learned to guard against preconceived notions and premature acceptance of conclusions. He has learned to reserve judgment and to modify and discard conclusions as fresh facts become available. In his service functions, as in his research, the applied psychologist often must formulate hypotheses and verify them against

subsequent data. In trying to discover the sources of a client's neurotic fears, the causes of a child's reading difficulties, or the leadership potential of a job applicant, the psychologist forms a succession of tentative hypotheses. Each hypothesis suggests what further facts must be elicited and is in turn either confirmed or disproved by such facts. This type of problem-solving attitude represents an important difference between the professional psychologist and both the well-intentioned but untrained "psychologizer" and the deliberate impostor.

**The problem of charlatanism.** Some of the areas in which applied psychologists work have traditionally proved attractive to charlatans and self-styled experts (see Anastasi 1964, pp. 10–13). In this flourishing psychological underworld are found phrenologists, physiognomists, and others who try to diagnose aptitudes and personality traits or recommend a specific career on the basis of irrelevant physical signs. Here too are the many pseudoscientific systems for improving personality, developing one's memory or selling ability, and overcoming emotional problems. On the borderline are the practitioners who assume a semblance of respectability by administering tests. They may, however, be relying on a single test of unknown validity for which they make extravagant claims. Or they may send a set of tests by mail, to be taken under uncontrolled conditions and interpreted with no other knowledge about the examinee. All these practices represent gross misuse of tests and yield meaningless results.

The prevalence of charlatanism in the field of human behavior arises partly from inadequate understanding of the nature of psychology. Since psychology is a relatively young science, the public is still unclear about the functions of psychologists and is thus more receptive to the claims of charlatans. Even more important in accounting for the survival of charlatanism is the public's strong desire for the sort of help offered by charlatans. Unhampered by facts, the charlatan promises easier and quicker solutions to life's problems than does the reputable psychologist. He either omits all mention of any margin of error in his procedures or gives a bland assurance of near-perfect accuracy. In support of his claims he is likely to cite an impressive array of selected cases, thereby capitalizing on chance results. Relying on analogies, figures of speech, and popular stereotypes, his writings create a spurious semblance of plausibility.

A favorite procedure of charlatans is to provide personality descriptions in terms of vague generalities applicable to most people. When confronted with such a description of himself, the individual



is often impressed with its apparent accuracy and insight and becomes convinced that the procedure is valid. This subjective pseudovalidation has been called the "Barnum effect" (Dunnette 1957) after Phineas T. Barnum, the famous showman, who is credited with remarking, "There's a sucker born every minute." The readiness with which these general personality descriptions are accepted as true has been demonstrated experimentally with such groups as personnel managers and college students (Stagner 1958). In these demonstrations a standard personality inventory is first administered to the group, and the papers are collected for scoring. Before getting the actual scores, each participant receives an identical copy of a fake "personality analysis." The analysis consists of a list of descriptive statements in which the same statements have been marked for each person. The marked statements, most of which are flattering, are such as to apply to the large majority of people. They are interspersed with unmarked statements, which are generally unfavorable. When asked to evaluate their self-descriptions, the majority of participants rate them as "amazingly accurate," while most of the others rate them as "rather good." Few if any disagree with them. Upon being told about the hoax and learning that their own description is identical with all the others, subjects typically show great surprise. The Barnum effect accounts for much of the popular acceptance of pseudoscientific systems of personality analysis. These systems provide random assortments of universally applicable generalities and fail to differentiate one person from another.

**Identifying applied psychologists.** The prevalence of charlatanism provides one reason for inquiring into ways of identifying genuine psychologists. In addition, the phenomenal growth of applied psychology since World War II means that more and more psychologists are working with laymen or with members of other professions who may lack the technical knowledge to appraise the psychologist's qualifications. Hence, it has become increasingly important to establish procedures for identifying psychologists and judging their special competencies.

A significant step, both in raising professional standards and in helping the general public to evaluate psychological practitioners, has been the enactment of licensing and certification laws for psychologists. Before World War II there was virtually no legal control over either the practice of psychology or the use of the title "psychologist." Today most of the states in which large numbers

of psychologists are employed either have such laws or are in the process of enacting them (Anastasi 1964, p. 584). In such states psychologists may be identified by the possession of a certificate indicating that certain standards of training and experience have been met. Typical requirements for certification include the PH.D. degree in psychology from an accredited university, two years of relevant experience, and the passing of a written qualifying examination.

At a still higher level, specialty certification is handled by the American Board of Examiners in Professional Psychology (ABEPP), established in 1946. Requiring the PH.D. degree, five years of experience, and intensive written and practical examinations, this board issues diplomas in three specialties: clinical, counseling, and industrial psychology. ABEPP publishes a directory of diplomates in each specialty, which is available in libraries and may also be obtained by interested organizations or individuals. The function of ABEPP is essentially to provide information about qualified psychologists. As a privately constituted board within the profession, it has no legal authority for enforcement. On the other hand, the various state laws pertaining to psychology do carry such authority. In a state having such laws a person without the appropriate certificate who presents himself as a psychologist is subject to legal penalties.

Another major source of information about psychologists is the American Psychological Association (APA), the national association to which most psychologists belong. Membership in this association does not imply qualifications for the performance of any service functions—only state certificates and ABEPP diplomas serve that purpose. Nevertheless, the annual directory of the APA provides considerable information about the location, training, and professional functions of its members. The APA has also undertaken many other activities designed to improve and systematize training standards, raise the level of professional practice, work out effective relations with other professions, and clarify the public image of psychology. It has participated in the preparation of two booklets designed to inform the public about the work of psychologists, *Psychologists in Action* (Ogg 1955) and *The Psychologist in Industry* (American Psychological Association, Division of Industrial Psychology 1959). Of particular importance was the formulation of a professional code of ethics, "Ethical Standards of Psychologists" (American Psychological Association 1963). First published in 1953 and revised in 1963, this code is

especially relevant to the problems encountered by applied psychologists in their relations with the public.

### Personnel psychology

One of the major areas of personnel psychology is that of *personnel selection and classification*. In industry the techniques of personnel selection and classification are used not only in hiring new employees but also in connection with promotions, transfers, discharges, and other personnel decisions. Nor are such techniques limited to industrial psychology. Essentially the same procedures are applicable to the selection and classification of military personnel, the admission of students to colleges and professional schools, and the identification of patients most likely to benefit from a particular form of therapy. Although first developed within the context of industrial psychology, these techniques provide an introduction to methodology that is of basic importance in most fields of applied psychology. [See INDUSTRIAL RELATIONS, *article on INDUSTRIAL AND BUSINESS PSYCHOLOGY*.]

A useful distinction is that between selection and classification of personnel. Selection always involves a decision to accept or reject. Decisions as to whether or not to hire a job applicant, admit a student to medical school, or accept a candidate for officer training are examples of selection. In classification, on the other hand, every individual must be utilized. The question is where to assign each person so as to maximize the total effectiveness of the organization. The development of a program for either selection or classification of personnel involves four principal steps: job analysis, preparation and preliminary tryout of predictors, validation of predictors, and formulation of a strategy for personnel decisions (Anastasi 1964, chapters 2-5).

Job analysis defines the problem by specifying worker qualifications for which predictors are needed. Through a variety of techniques for the systematic observation of workers on the job, a job description can be prepared that is specific to each job. A description in terms of vague generalities that would be equally applicable to many jobs is of little use for personnel selection or classification. Moreover, effective job analysis identifies critical requirements, that is, those characteristics that differentiate most sharply between better and poorer workers on each job.

On the basis of the job analysis suitable predictors are chosen or developed. These include not only tests but also other assessment procedures,

such as application forms, reference checkups, and interviewing. Application forms (or biographical inventories) have proved successful in a wide variety of selection situations, particularly when their items are empirically chosen and weighted by checking them against criteria of subsequent job performance. Although more time-consuming than other techniques, intensive interviewing provides opportunities for thorough individual assessment. Interviewing techniques are employed especially in the evaluation of executives and other high-level personnel and in the judging of many personality traits for which no satisfactory tests may be available. At a more basic level considerable research has been conducted on the development of rating scales and other systematic procedures for improving personal judgments (Anastasi 1964, pp. 76-88). [See INTERVIEWING.]

Every major type of test has been used in personnel selection and classification (Anastasi 1961). Traditional intelligence tests measure principally verbal comprehension and other abstract functions required for academic achievement. Several short intelligence tests have been developed for screening industrial personnel. Research on creative talent has produced many new testing techniques that are relevant to certain selection problems. Among the earliest tests applied to personnel selection are tests of special aptitudes in motor skills, spatial aptitudes, mechanical comprehension, and clerical speed and accuracy. Classification programs are making increasing use of multiple aptitude batteries, which combine the traits measured by traditional intelligence tests with some of the broader abilities covered by special aptitude tests. Achievement tests measure current proficiency following relatively uniform training or experience. Many consist of standardized work samples, such as taking dictation, operating a punch press, or handling the matters in an executive's in-basket. Personality tests may be used with any type of worker but have found their chief application in jobs requiring extensive interpersonal contacts, such as selling and supervisory positions. It is especially important in this area to guard against common misuses of tests, including undue reliance on test scores in the absence of other information and the use of inadequately validated instruments. Among the most successful personality tests are interest inventories, whereby the individual may compare his interests with those typical of persons engaged in different occupations. [See APTITUDE TESTING; INTELLIGENCE AND INTELLIGENCE TESTING; PERSONALITY MEASUREMENT; VOCATIONAL INTEREST TESTING.]



Prior to their operational use all predictors should be validated within a representative sample of the population for whom the selection program is designed. The most common procedure is to correlate predictive measures (e.g., test scores, interviewer's ratings) with a criterion of subsequent job success. On the basis of such predictive validity some of the instruments will be retained and others discarded or modified. The validation data also permit an evaluation of the effectiveness of the final selection program and a determination of the margin of error to be expected in its operation. [See PSYCHOMETRICS.]

The final stage in the development of the program requires the choice of a strategy for applying the predictive measures so as to reach a decision about individuals (Cronbach & Gleser 1957). Scores on the different predictors may be combined through a regression equation, which yields the single best estimate of each individual's criterion performance. Through such a procedure deficiencies in one qualification may be compensated by superior standing in another. In jobs where minimum standards must be met in each of several essential skills, a multiple-cut-off strategy is more appropriate. When tests or other predictors are employed in the intensive study of individual cases, as in the selection of high-level personnel, evaluation of executives, and clinical diagnosis, it is a common practice for the examiner or interviewer to combine and interpret scores through "clinical judgment," with no further statistical analysis (Meehl 1954).

A growing area within personnel psychology is that of *development and training* (Anastasi 1964, chapter 5; Gagné 1962, chapters 9-11). Although the psychology of learning has much to contribute to the improvement of training procedures, well-designed applied research in this area is still rare. Industrial training is a continuing process that can benefit all employees, from beginners to experienced workers and from unskilled labor to top management. Training programs in industry serve a variety of functions, including orientation, development of job skills, technical and professional training, management development, and general education.

The development of a training program should include task analysis to define what must be learned, development of appropriate training procedures, and objective evaluation of the program. The choice of training media and procedures depends in part upon what is to be taught and to whom. In developing training procedures, one can

also be guided by the results of available training research in connection with such questions as motivation, stress, practice, feedback, and transfer of training. The effectiveness of a training program should be evaluated in terms of job criteria as well as training criteria.

Recent developments in the automation of training procedures are illustrated by training devices, simulators, and teaching machines. The rapidly growing area of management development covers many types of training and utilizes a diversity of training techniques. Much of it is concerned with the development of human-relations skills and attitudes through group discussion of specific cases, problem-solving interviews, and role playing (Maier, Solem, & Maier 1957). In other aspects of management development two techniques of special interest are management games (or simulation) and creativity training through "brainstorming." The former provides practice in executive decision making, often utilizing computers to furnish rapid feedback on the probable consequences of decisions (Fleishman 1961, pp. 219-238). The latter endeavors to stimulate creative productivity by requiring participants to separate in time the production of ideas from their evaluation (Parnes & Meadow 1963). It is hypothesized that the individual is thus freed from the customary inhibitory effects of self-criticism.

A third and relatively broad area of personnel psychology is concerned with *employee motivation and morale* (Anastasi 1964, chapter 6; Fleishman 1961, secs. 4-6). Various designated as "industrial social psychology," "employee relations," "human relations in industry," and "the psychology of management," this division of personnel psychology dates from the classic Hawthorne studies, which first focused attention on the effect of employee attitudes upon productivity (Roethlisberger & Dickson 1939). Incidentally, the Hawthorne studies also served to highlight the need for controlling attitudinal variables in the design of psychological experiments. The term "Hawthorne effect" is now commonly used to designate the influence that participating in an experiment may have upon the subject's behavior.

The principal procedures used in gathering data on employee motivation, attitudes, and morale include interviews, opinion questionnaires, attitude scales, projective and other indirect techniques, nominating or sociometric techniques, and continued observation of small groups by a trained observer. When rating the factors that make for job satisfaction, workers usually place as high a

value on security, opportunity for advancement, interest in their work, feeling of pride in the company, and relations with co-workers and supervisors as they do on pay. However, the relative weight given each factor varies somewhat with the age, sex, and education of the worker, as well as with type of work and job level. Favorable employee attitudes are associated with a lower accident rate and with less absenteeism, tardiness, and job turnover. The relation between attitudes and productivity, however, is more complex and depends upon other individual and situational variables. There is also evidence to suggest that favorable job attitudes more often arise from intrinsic characteristics of the work itself, while unfavorable job attitudes more often arise from the job context (Herzberg, Mausner, & Snyderman 1959). Moreover, favorable attitudes bear a more direct relation to productivity than do unfavorable ones.

The attitudes and morale of any group are deeply influenced by the supervisory practices, the communication system, and the organizational structure under which it operates (Likert 1961). Research on supervisory and leadership behavior has identified two independent dimensions: "consideration," characterized by a friendly, warm, considerate supervisory relationship, as contrasted with an impersonal and authoritarian relationship; and "initiating structure," pertaining to the extent to which the supervisor actively plans and directs group activities oriented toward goal attainment (Fleishman 1961, pp. 311-328). Both dimensions are related to employee attitudes and productivity.

Research on industrial communication has covered such varied problems as the relative merits of different media; legibility, content, and readability of written communications; the development of communication skills; factors influencing the persuasiveness of communication; sources of communication barriers and breakdowns; and the nature and effectiveness of communication networks (Fleishman 1961, sec. 6).

Organizational structures have been analyzed from many points of view, and this diversity of approach is reflected in the proposals for improving organizational effectiveness (Foundation for Research on Human Behavior 1959; Bass 1965). Although controlled experiments pertaining to organizational theory are few, some research is available on the effectiveness of such organizational practices as employee participation in decision making, sociometric work-group assignments, and the redesigning of jobs with reference to attitudinal factors (Fleishman 1961, secs. 4-6). While all

these approaches are promising, it must be borne in mind that no single organizational formula, management procedure, or supervisory practice fits all jobs and all individuals.

### Engineering psychology

In current usage *engineering psychology* is a rather flexible term. This flexibility also characterizes its various synonyms—human engineering, human-factors engineering, and ergonomics. All these terms are sometimes used in a broad sense to cover practically every topic in industrial psychology, including selection, training, motivation, work methods, working environment, and equipment design. A more common usage restricts the coverage to the last three topics, as is done in this article. It should also be noted that the techniques of engineering psychology extend beyond industrial psychology and are being applied more and more to the designing of consumer goods. Psychologists are contributing not only to the designing of industrial machinery, military equipment, and space vehicles but also to the designing of such products as kitchen stoves, radios, and telephone dials. [See ENGINEERING PSYCHOLOGY.]

A major objective in the improvement of work methods is the reduction of *fatigue* (Anastasi 1964, chapter 7; Barnes 1958, chapter 14; Fleishman 1961, sec. 7). Fatigue has been variously defined in terms of feelings of tiredness, changes in output, and physiological conditions. Increase in the energy cost of work, as measured by the respiration calorimeter, illustrates the physiological concept of fatigue. Vigilance tests provide a promising behavioral index of fatigue. Feelings of monotony and boredom are difficult to differentiate from the subjective experience of fatigue, especially from so-called mental fatigue.

Human efficiency can be increased through the design of *work and rest schedules*. Suitably scheduled rest pauses generally increase total output, although the improvement may not occur immediately. Lengthening the work week beyond an optimal number of hours has usually lowered rather than raised total output. Shift work involves a disruption of daily physiological rhythms, especially for workers on rotating shifts. Individuals differ widely in their ability to adjust to changing activity cycles.

Beginning with the early contributions of Frederick W. Taylor and Frank B. Gilbreth, *time-and-motion study* has now become part of more comprehensive methods-improvement programs (Barnes 1958). The principal objective of time-and-motion



study is the elimination of unnecessary movements in work processes. Modern methods-improvement programs are broader in scope than early time-and-motion study, and they give more attention to individual differences and worker participation in methods development.

*Accident prevention* is a recognized objective of modern methods improvement, besides being related to nearly every other aspect of industrial psychology (Anastasi 1964, chapter 7; Fleishman 1961, sec. 8). Data on the relation of personal characteristics to accident rate are often difficult to interpret because of the interrelation of many personal variables. Accident-proneness can be conclusively demonstrated only through the comparison of individual accident records over different time periods. Studies of "psychological climate" in different industrial settings have shown accident rates to be significantly related to several situational variables. With the recognition of the importance of such variables as causes of traffic accidents, the methods of engineering psychology are being used increasingly to improve the layout, illumination, and marking of highways. Accidents can be regarded as one indicator of the deterioration of performance under adverse conditions.

It has long been known that working efficiency depends in part upon the characteristics of the working environment. It is apparent that such problems pertain not only to factories and offices but also to schools, libraries, homes, waiting rooms, and any other environments designed for prolonged human use. From a different angle, recent developments in space travel have stimulated intensive research on the effects of certain extreme environmental conditions upon human performance.

With regard to *illumination*, research has concentrated on the optimum intensity of light for different types of work, the distribution of light to minimize glare and other conditions producing visual fatigue, and the use of different colors of light (McCormick 1964, chapter 13). The primary object of *ventilation* is to enable the body to give off excess heat through evaporation, radiation, and convection (McCormick 1964, chapter 15). Industrial studies have shown significant effects of ventilating conditions upon the quantity and quality of production, accident rates, and absenteeism. Suggestive results on the behavioral effects of weather, seasons, and climate were obtained in a number of early studies (Anastasi 1964, pp. 203-204). Recently there has been a revival of interest in these problems, known collectively as "psychoclimatology." Research is also being conducted on the psychological consequences of exposure to ex-

treme atmospheric conditions, among which oxygen deprivation has received the greatest attention from psychologists.

Chief among the special psychological problems of *space travel* are those arising from extreme and stressful environmental conditions (McCormick 1964, chapter 16). The physical conditions investigated in this connection include weightlessness, rotation, high accelerative forces, and conflicting sensory cues leading to spatial disorientation. A major source of psychological stress is isolation, comprising the factors of confinement, detachment, and sensory deprivation. [See SPACE, OUTER, article on SOCIAL AND PSYCHOLOGICAL ASPECTS.]

Noise has been investigated as a cause of temporary or permanent hearing impairment, as a source of interference with auditory communication, and as a major form of distraction (McCormick 1964, chapter 14; Fleishman 1961, pp. 512-519). The effects of noise distraction upon behavior depend upon the interaction of various characteristics of the distracting stimuli, the task, and the individual. Their proper interpretation requires data not only on output but also on energy expenditure, adaptation, and emotional response to the source of distraction. A suggestive hypothesis, for which some experimental evidence is available, is that performance is most effective when the subject is at an optimum level of arousal (McBain 1961). Thus the introduction of a varying noise should improve the performance of an intrinsically monotonous task, since it brings the initially low arousal level closer to the optimum.

Although foreshadowed by the contributions of Taylor, Gilbreth, and other early investigators, research on human factors in *equipment design* dates largely from World War II (Anastasi 1964, chapter 9; Barnes 1958, chapters 16-17; McCormick 1964, chapters 5-12). As the complexity of machines began to outstrip man's capacity to operate them, it became necessary to redesign equipment in such ways as to make the operator's task less demanding. One of the problems of equipment design concerns the spatial layout of work. This is illustrated by the pre-positioning of pens, screwdrivers, and other small tools in holders and by the semicircular work space within which all materials or controls are easily reached by the operator's hands. Another problem pertains to the improvement of displays through which information is transmitted to the operator. For example, extensive research has been conducted on the factors influencing speed and accuracy of dial reading, such as dial size and shape, length of scale units, number and spacing of scale markers, direction of pointer movements,

and the size and style of letters and numerals. Research on control problems is illustrated by the hundreds of experiments on tracking. Another example is provided by research on the coding of controls in terms of position, size, shape, or color. Confusion among controls can be reduced by standardizing each control, increasing the discriminability of different controls, and utilizing meaningful associations for their identification.

The broadening scope of engineering psychology is best typified by its growing concern with *man-machine systems* (Fleishman 1961, sec. 9; Gagné 1962; McCormick 1964, chapters 17-19). By describing the functions of the human operator in engineering terminology, such an approach makes it possible to transfer some of these functions from the man to the machine. The systems approach also permits the efficient arrangement of men and machines within a complex system. It is through the study of man-machine systems that the greatest improvements in speed and accuracy of operation have been achieved. It is also through this approach that engineering psychology is most likely to contribute to the advancement of basic psychological knowledge.

### Consumer psychology

While personnel and engineering psychology are concerned chiefly with the behavior of the individual as producer, consumer psychology deals with his behavior as consumer (Anastasi 1964, chapters 10-12). This field of psychology began with the psychology of advertising and selling, whose objective was effective communication from the manufacturer or distributor to the consumer. Through advertising the consumer is informed about available goods or services and about the ways in which they may meet his needs. More recently, consumer psychology has broadened to cover two-way communication—from consumer to producer as well as from producer to consumer. Communications from the consumer were first sought as a means of preparing more effective advertisements. Systematic surveys of consumer needs and wants enabled the advertiser to identify those features of his product that were of greatest interest to the consumer. Such consumer surveys led directly to the next step, which involved the consideration of consumer wants and preferences in the designing of products.

The methods of consumer psychology are applicable to a number of fields other than that of developing and marketing industrial goods and services. One such application is to be found in public opinion polls, whose procedures are essen-

tially the same as those of consumer surveys. Another example is provided by the well-designed research on food preferences conducted by the armed services as an aid in the planning of meals for military personnel. A final illustration is the use of advertising campaigns in connection with problems of public interest, ranging from the prevention of forest fires to the promotion of mental health. [See CONSUMERS, *article on* CONSUMER BEHAVIOR; MARKET RESEARCH, *article on* CONSUMER RESEARCH.]

Modern consumer psychology covers a wide variety of problems, and its methods are correspondingly diverse. One of its most active areas deals with procedures for testing the effectiveness of printed advertisements, such as readership surveys, analysis of coupon returns, and controlled surveys of brand use and sales records. Specialized techniques have been developed for measuring the extent to which an ad attracts and holds attention, is understood and recalled, arouses appropriate associations and feeling tones, and leads to the desired action. Through controlled field or laboratory investigations, the effectiveness of ads in all these respects has been related to such ad characteristics as size, color, amount of white space, illustrations, printing type, content of the written message, position in the medium, and number of repetitions. Research on trade names has dealt with such questions as ease of identification and recall, arousal of appropriate feeling tone, and avoidance of confusion with other trade names.

In *radio and television research* a variety of specialized techniques has likewise been employed to gather data on audience size and composition, audience reaction to specific programs or to types of programs, and sales effectiveness of advertising.

A flurry of public agitation regarding *subliminal advertising* has proved to be largely unfounded (McConnell, Cutler, & McNeil 1958). Subliminal perception is the process of responding to stimuli that fall below the threshold of verbally reported awareness. The results of laboratory experiments indicate that subliminal stimulation cannot appreciably affect the behavior of mass audiences, contrary to the claims of certain popular writers and of the promoters of this type of advertising.

Cutting across several areas of consumer psychology, *motivation research* utilizes depth interviewing and projective techniques to study buying motives. The principal contribution of this approach lies in providing hypotheses for subsequent testing by more objective, quantitative methods applicable to large samples. Information regarding consumer motivation can also be obtained by more direct



forms of inquiry if suitable experimental designs are employed. The "product image," which refers to the composite of ideas and feelings associated with a product, has been investigated by the methods of motivation research as well as by more controlled techniques such as adjective checklists and the semantic differential. [See ADVERTISING, *article on ADVERTISING RESEARCH.*]

*Personal salesmanship* may be regarded as the most flexible and individualized form of advertising. Because of the many varieties of selling, generalizations about effective salesmanship are likely to be of little value. Psychologists have concentrated on procedures for the selection and training of salesmen for specific jobs. Available research on the sales process is still relatively meager.

### Clinical psychology

Characterized by intensive work with individual cases, clinical psychology is concerned chiefly with intellectual and emotional disorders (Anastasi 1964, chapters 13-15; Shaffer & Shoben 1956). Today clinical psychologists work in a variety of settings, including not only hospitals and outpatient clinics but also industry, schools, prisons, courts, government agencies, and the armed services. The practice of clinical psychology utilizes findings and techniques from many branches of psychology, such as abnormal psychology, personality theory, and psychological testing. The specific activities of clinical psychologists consist chiefly of diagnosis, therapy, and research, although the relative emphasis placed upon each varies widely with the work setting and with the individual psychologist. Moreover, these three functions are frequently intertwined. Research may be specifically concerned with the validity of diagnostic tools or with the effectiveness of different types of therapy. Similarly, diagnostic and therapeutic procedures are indistinguishable in certain types of psychotherapy that emphasize the therapeutic value of the client's insight into the nature and origin of his problems. [See CLINICAL PSYCHOLOGY.]

In its broadest sense, *diagnosis* corresponds to the fact-finding aspect of clinical practice. Its objectives include screening and classification, personality description, prediction of outcome, and attainment of insight by the client. In the classification of psychological disorders a basic distinction is that between mental deficiency and emotional problems, the latter being subdivided into psychoses and neuroses. More specific syndromes have been identified, but the application of these traditional psychiatric categories has many drawbacks. Characteristically the clinical psychologist is inter-

ested in a detailed personality description of the individual case, with its unique combination of problems, adjustment mechanisms, and antecedent circumstances.

The major sources of data available to the clinician include the case history, psychological testing, and the diagnostic interview. Considerable discussion and some research have been devoted to the nature and effectiveness of "clinical judgment" (Meehl 1954). With regard to both data gathering and interpretation, there are a number of ways in which the services of a skilled clinician are needed. Whether objective diagnostic techniques and statistical prediction formulas can eventually assume the entire diagnostic function remains a moot point.

Many kinds of *therapy* are used today in treating emotional disorders. The choice of therapy depends partly upon the nature and severity of the disorder and partly upon such client characteristics as age and intelligence. To a considerable extent, however, the type of therapy employed reflects the training and theoretical persuasion of the therapist, available facilities, and other extraneous factors. The same behavioral symptoms may be treated quite differently by different therapists; and several types of therapy may be combined in treating a single individual. There are few forms of treatment that are wholly specific to a given emotional disorder. In psychology, as contrasted to medicine, one rarely if ever finds a disorder in which the diagnosis completely determines the treatment. Evaluating the relative efficacy of different forms of therapy for emotional disorders is very difficult. Owing to methodological difficulties, there is as yet little evidence to demonstrate that one therapy is better than another for any given case or even that any therapy is effective at all. Since the practical demands of the situation call for some action, the therapist does the best he can within the limits of available knowledge.

Available therapies for emotional disorders include somatic therapies, environmental therapies, psychotherapy, and preventive procedures. Somatic therapies, requiring the services of a licensed physician, are illustrated by psychosurgery, shock therapies, and various applications of psychopharmacology. The introduction of tranquilizers has led to a spectacular rise in the recovery rate of patients in American state mental hospitals. A major factor in these results appears to be the role of tranquilizers in making patients more amenable to psychotherapy and, indirectly, improving staff morale. Other, more direct effects of such drugs have not yet been conclusively established in adequately con-

trolled experiments. Environmental therapy includes the therapeutic manipulation of environmental conditions both in community contexts (e.g., home, school, job, or recreational centers) and in institutions. A growing emphasis upon milieu therapy is rapidly transforming mental hospitals from custodial institutions into therapeutic communities. [See MENTAL DISORDERS, TREATMENT OF, *article on SOMATIC TREATMENT.*]

Psychotherapy is practiced in many forms, reflecting the diverse theoretical orientations of the therapists. In the most general terms, psychotherapy can be described as the use of verbal or other means of communication in a nonthreatening interpersonal situation that facilitates the alleviation of anxiety and the elimination of maladaptive behavior through a learning process. Two widely used types of psychotherapy are psychoanalysis (Dollard & Miller 1950) and client-centered therapy (Rogers 1951). Group therapy is increasingly being employed, either in place of individual psychotherapy or in conjunction with it. Psychodrama and role playing are special variants of group therapy. Play therapy is commonly utilized with children, in lieu of the predominantly verbal techniques employed with adults. The current mental health movement is concerned not only with the treatment and rehabilitation of institutionalized mental patients but also with preventive measures and with the fostering of positive mental health through comprehensive social and educational programs. [See MENTAL DISORDERS, TREATMENT OF, *articles on CLIENT-CENTERED COUNSELING and GROUP PSYCHOTHERAPY.*]

It is in research that the clinical psychologist, as contrasted with other mental health professionals, can make his most distinctive contribution (Conference on Research in Psychotherapy 1962). Even in his service functions, the clinical psychologist with a research orientation can operate more effectively than one who is trained only as a practitioner. With research training he is better able to evaluate innovations and incorporate new developments into his practice. In a rapidly evolving field rigid commitment to existing techniques can only retard progress.

A major area of applied research in clinical psychology deals with the validity of diagnostic procedures, particularly projective techniques and other special clinical instruments. Research on the process of psychotherapy has been concerned with changes in client behavior in the course of therapy, the specific ways in which different therapists function, and the interactions among therapist, client, and process variables (Rogers & Dymond 1954).

Special techniques have been developed to measure changes in the client's self concept in the course of therapy. The outcome of therapy is particularly difficult to measure satisfactorily. Among the criteria of outcome employed for this purpose are client reports, the therapist's evaluation of progress, objective measures of behavioral changes during therapy sessions, ward behavior of institutionalized cases, and follow-up data in real-life situations.

Research on the causes of mental disorders is an interdisciplinary venture spanning many fields, from biochemistry to anthropology. The sources of intellectual and emotional disorders have been sought in heredity, environmentally produced organic conditions, and experiential factors. Basic research that integrates clinical and experimental approaches is beginning to make vital contributions both to clinical practice and to the science of psychology. The reformulation of clinical concepts in terms of learning theory has been especially productive of research that cuts across clinic and laboratory lines (Dollard & Miller 1950; Grossberg 1964).

### Counseling psychology

While there is much overlapping between the work of clinical and counseling psychologists, the latter typically deal with relatively normal persons having minor emotional problems or wanting help in vocational, educational, or other practical decisions (Anastasi 1964, chapter 16; Bordin 1955; Tyler 1961). Like psychotherapy, counseling relies largely on verbal communication and an accepting interpersonal relationship. In contrast to psychotherapy, however, counseling is less concerned with changing the individual's basic personality structure and more concerned with the effective utilization of his present resources. Counseling is also likely to be of shorter duration than psychotherapy and to involve the giving of more factual information to the client. Counseling psychologists work in many settings and perform a variety of functions. Although they may specialize in one type of counseling, such as vocational, educational, employee, rehabilitation, old-age, marriage, or personal counseling, they recognize the interrelation of all adjustment problems and concentrate on counseling the whole person. [See COUNSELING PSYCHOLOGY.]

To increase the client's self-knowledge and assist him in making decisions, counseling psychologists utilize several types of tests, notably intelligence tests, educational achievement tests, special aptitude tests, multiple aptitude batteries, interest inventories, and personality inventories. Because of



the importance of vocational decisions in most counseling activities, counseling psychologists need to make current occupational information available to their clients. Both tests and occupational data should be integrated in the counseling process in such a way that the client accepts the information and uses it effectively (Bordin 1955; Goldman 1961; Tyler 1961). An important objective of counseling is to facilitate decision making, not only by providing self-knowledge and occupational or other relevant information but also through training in the decision process.

One of the principal areas of counseling research concerns the process of occupational choice and the nature of vocational development throughout the life span (Anastasi 1964, chapter 17; Roe 1956; Super 1957; Super & Overstreet 1960). In selecting an occupation, the individual chooses a way of life. One's occupation is likely to affect not only his income level and general social status but also many specific aspects of daily life. Job satisfaction and general mental health are contingent upon the extent to which the individual's work satisfies needs and values that are significant to him. Considerable information is available on personality patterns characteristic of persons in different occupations, but the results are often difficult to interpret because of procedural inadequacies. Some research has also been done on the role of childhood experiences in vocational choice. [See VOCATIONAL INTEREST TESTING.]

### Psychology and other professions

Chief among the fields in which psychologists work in consulting or collaborative relations with other professional personnel are education, medicine, and law (Anastasi 1964, chapters 18–20). In education, psychologists may serve as school psychologists or as educational psychologists. The functions of school psychologists consist largely in the diagnostic testing and interviewing of individual pupils, followed by recommendations for remedial action. Educational psychologists are concerned principally with teacher training and educational research. While their chief focus is on the psychology of learning, they also draw upon other areas, notably developmental psychology, differential psychology, psychological testing, the psychology of adjustment, and social psychology. [See EDUCATIONAL PSYCHOLOGY.]

The potential contributions of psychology to medicine are still largely undeveloped. Although psychological techniques and knowledge can aid in every stage of medical practice, including diagnosis, treatment, rehabilitation, and prevention, contacts between psychology and general medicine

prior to World War II were negligible. Since mid-century the participation of psychologists in medical school teaching has grown rapidly. There have been corresponding increases in the utilization of psychologists as members of interprofessional teams, as for example, in the rehabilitation of the physically handicapped. Psychologists are likewise engaging more and more in collaborative research projects conducted in medical settings. [See VOCATIONAL REHABILITATION.]

The earliest applications of psychology to law were concerned with the evaluation of testimony and other courtroom procedures. A related area is that of lie detection. In an effort to identify the causes of antisocial behavior, psychologists have studied the characteristics of delinquents and criminals. Although the conditions that lead to crime are manifold, research findings give increasing prominence to emotional and motivational factors stemming, at least in part, from childhood experiences. The work of psychologists in court clinics, prisons, training schools for juvenile delinquents, and other penocorrectional settings falls chiefly within the fields of clinical and counseling psychology. At a more basic level, psychologists may aid in the development, revision, and interpretation of laws by submitting relevant research findings in court testimony, by serving as consultants to legislative committees, and by collaborating in law school teaching and research. Empirically established facts about human behavior are basic to many types of laws, such as those pertaining to insanity, drug addiction, alcoholism, sexual offenses, child development, education, and minority groups.

ANNE ANASTASI

[Directly related are the entries ADVERTISING, article on ADVERTISING RESEARCH; CLINICAL PSYCHOLOGY; CONSUMERS, article on CONSUMER BEHAVIOR; COUNSELING PSYCHOLOGY; EDUCATIONAL PSYCHOLOGY; ENGINEERING PSYCHOLOGY; INDUSTRIAL RELATIONS, article on INDUSTRIAL AND BUSINESS PSYCHOLOGY; MARKET RESEARCH, article on CONSUMER RESEARCH; MILITARY PSYCHOLOGY. Other relevant material may be found in APTITUDE TESTING; FATIGUE; LEARNING, article on ACQUISITION OF SKILL.]

### BIBLIOGRAPHY

- AMERICAN PSYCHOLOGICAL ASSOCIATION 1963 *Ethical Standards of Psychologists*. *American Psychologist* 18:56–60.
- AMERICAN PSYCHOLOGICAL ASSOCIATION, DIVISION OF INDUSTRIAL PSYCHOLOGY 1959 *The Psychologist in Industry*. Washington: Research Institute of America.
- ANASTASI, ANNE 1961 *Psychological Testing*. 2d ed. New York: Macmillan. → The first edition was published in 1954.
- ANASTASI, ANNE 1964 *Fields of Applied Psychology*. New York: McGraw-Hill.

- BARNES, RALPH M. 1958 *Motion and Time Study*. 4th ed. New York: Wiley. → First published in 1942 as *Motion and Time Study Applications*.
- BASS, BERNARD M. 1965 *Organizational Psychology*. Boston: Allyn & Bacon.
- BORDIN, EDWARD S. 1955 *Psychological Counseling*. New York: Appleton.
- CONFERENCE ON RESEARCH IN PSYCHOTHERAPY, 1961 1962 *Research in Psychotherapy: Proceedings*. Edited by Hans H. Strupp and L. Luborsky. Washington: American Psychological Association.
- CRONBACH, LEE J.; and GLESER, GOLDINE C. (1957) 1965 *Psychological Tests and Personnel Decisions*. Urbana: Univ. of Illinois Press.
- DOLLARD, JOHN; and MILLER, NEAL E. 1950 *Personality and Psychotherapy: An Analysis in Terms of Learning, Thinking and Culture*. New York: McGraw-Hill. → A paperback edition was published in 1965.
- DUNNETTE, M. D. 1957 Use of the Sugar Pill by Industrial Psychologists. *American Psychologist* 12:223-225.
- FLEISHMAN, EDWIN A. (editor) 1961 *Studies in Personnel and Industrial Psychology*. Homewood, Ill.: Dorsey.
- FOUNDATION FOR RESEARCH ON HUMAN BEHAVIOR 1959 *Modern Organization Theory: A Symposium*. Edited by Mason Haire. New York: Wiley.
- GAGNÉ, ROBERT M. (editor) 1962 *Psychological Principles in System Development*. New York: Holt.
- GOLDMAN, LEO 1961 *Using Tests in Counseling*. New York: Appleton.
- GROSSBERG, JOHN M. 1964 Behavior Therapy: A Review. *Psychological Bulletin* 62:73-88.
- HERZBERG, FREDERICK; MAUSNER, BERNARD; and SNYDERMAN, BARBARA BLOCH 1959 *The Motivation to Work*. 2d ed. New York: Wiley. → The first edition was published in 1957.
- LIKERT, RENSIS 1961 *New Patterns of Management*. New York: McGraw-Hill.
- MCBAIN, WILLIAM N. 1961 Noise, the "Arousal Hypothesis," and Monotonous Work. *Journal of Applied Psychology* 45:309-317.
- MCCONNELL, JAMES V.; CUTLER, RICHARD L.; and MCNEIL, ELTON B. 1958 Subliminal Stimulation: An Overview. *American Psychologist* 13:229-242.
- MCCORMICK, ERNEST J. 1964 *Human Factors Engineering*. 2d ed. New York: McGraw-Hill. → First published in 1957 as *Human Engineering*.
- MAIER, NORMAN R. F.; SOLEM, ALLEN R.; and MAIER, AYESHA A. 1957 *Supervisory and Executive Development: A Manual for Role Playing*. New York: Wiley.
- MEEHL, PAUL E. 1954 *Clinical Versus Statistical Prediction*. Minneapolis: Univ. of Minnesota Press.
- OGG, ELIZABETH 1955 *Psychologists in Action*. Washington: Public Affairs Committee.
- PARNES, SIDNEY J.; and MEADOW, ARNOLD 1963 Development of Individual Creative Talent. Pages 311-320 in *Research Conference on the Identification of Creative Scientific Talent, Scientific Creativity: Its Recognition and Development*. Edited by Calvin W. Taylor and Frank Barron. New York: Wiley.
- ROE, ANNE 1956 *The Psychology of Occupations*. New York: Wiley.
- ROETHLISBERGER, FRITZ J.; and DICKSON, WILLIAM J. (1939) 1961 *Management and the Worker: An Account of a Research Program Conducted by the Western Electric Company, Hawthorne Works, Chicago*. Cambridge, Mass.: Harvard Univ. Press. → A paperback edition was published in 1964 by Wiley.
- ROGERS, CARL R. 1951 *Client-centered Therapy: Its Current Practice, Implications, and Theory*. Boston: Houghton Mifflin.
- ROGERS, CARL R.; and DYMOND, ROSALIND F. (editors) 1954 *Psychotherapy and Personality Change: Coordinated Research Studies in the Client-centered Approach*. Univ. of Chicago Press.
- SHAFFER, LAURANCE F.; and SHOBEN, EDWARD J. JR. 1956 *The Psychology of Adjustment*. 2d ed. Boston: Houghton Mifflin. → The first edition was published in 1936.
- STAGNER, ROSS 1958 The Gullibility of Personnel Managers. *Personnel Psychology* 11:347-352.
- SUPER, DONALD E. 1957 *Psychology of Careers*. New York: Harper.
- SUPER, DONALD E.; and OVERSTREET, PHOEBE L. 1960 *Vocational Maturity of Ninth Grade Boys*. Horace Mann-Lincoln Institute of School Experimentation, Career Pattern Study, Monograph 2. New York: Columbia Univ., Teachers College.
- TYLER, LEONA E. 1961 *The Work of the Counselor*. 2d ed. New York: Appleton. → The first edition was published in 1953.

## PSYCHOMETRICS

*Psychometrics, broadly defined, includes all aspects of the science of measurement of psychological variables and all research methodologies related to them. In addition to this article, the area of psychometrics is also discussed in ACHIEVEMENT TESTING; APTITUDE TESTING; EXPERIMENTAL DESIGN; FACTOR ANALYSIS; INTELLIGENCE AND INTELLIGENCE TESTING; LATENT STRUCTURE; PERSONALITY MEASUREMENT; PSYCHOPHYSICS; QUANTAL RESPONSE; RESPONSE SETS; SCALING; SOCIOMETRY; VOCATIONAL INTEREST TESTING. Also of relevance are ATTITUDES and PROJECTIVE METHODS.*

*This article deals with five major topics: measurement, item analysis, reliability, validity, and test norms.*

### Measurement

Measurement is generally considered to be any procedure whereby numbers are assigned to individuals (used herein to mean persons, objects, or events) according to some rule. The rule usually specifies the categories of an attribute or some quantitative aspect of an observation, and hence defines a *scale*. A scale is possible whenever there exists a one-to-one relationship between some of the properties of a group of numbers and a set of operations (the measurement procedure) which can be performed on or observed in the individuals. Scales of measurement are commonly classified as nominal scales, ordinal scales, interval scales, and ratio scales; the variables they measure can be discrete (i.e., providing distinct categories that vary from each other in a perceptibly finite way) or continuous (i.e., not readily providing dis-



tinct categories; varying by virtually imperceptible degrees).

**Nominal scales.** In a nominal scale the numbers merely identify individuals or the categories of some attribute by which individuals can be classified. Letters or words or arbitrary symbols would do just as well. Simple identification is illustrated by the assignment of numbers to football players; classification by assigning numbers to such attributes as sex, occupation, national origin, or color of hair. We can cross-classify according to the categories of two or more attributes; e.g., sex by occupation, or sex by occupation by national origin. With nominal scales that classify, the variables are always treated as discrete. Sex, occupation, and national origin are genuine discrete variables. Color of hair, on the other hand, is a multidimensional continuous variable. If, for instance, we treat it as a discrete variable by establishing the categories blond, brunette, and redhead, the measure becomes unreliable to some degree, and some individuals will be misclassified. Where such misclassification can occur, the scale may be termed a *quasi-nominal* scale. Subject to the limitation of unreliability, it has the properties and uses of any other nominal scale.

The basic statistics used with nominal scales are the numbers, percentages, or proportions of individuals in the categories of an attribute or in the cells of a table of cross classification. Hypotheses about the distribution of individuals within the categories of one attribute, or about the association of attributes and categories in a table of cross classification, are usually tested with the chi-square test. Descriptive statistics used include the mode (the category which includes the largest number of individuals) and various measures of association, the commonest of which is the contingency coefficient. These statistics remain invariant when the order of the categories of each attribute is rearranged and when the numbers that identify the categories are changed. If the categories of an attribute have a natural order, this order is irrelevant to nominal scaling, and nominal-scale statistics do not use the information supplied by any such order. [See COUNTED DATA; STATISTICS, DESCRIPTIVE, article on ASSOCIATION; SURVEY ANALYSIS, article on THE ANALYSIS OF ATTRIBUTE DATA.]

**Ordinal scales.** An ordinal scale is defined by a set of postulates. We first introduce the symbol " $>$ ," and define it broadly as almost any asymmetrical relation; it may mean "greater than," "follows," "older than," "scratches," "pecks," "ancestor of," etc. We also define " $\neq$ " to mean "is unequal to" or "is different from." Then, given any class of

elements (say,  $a, b, c, d, \dots$ ) the relation  $>$  must obey the following postulates:

If  $a \neq b$ , then  $a > b$  or  $b > a$ .

If  $a > b$ , then  $a \neq b$ .

If  $a > b$  and  $b > c$ , then  $a > c$ .

If  $a, b, c, \dots$ , are the positive integers and  $>$  means "greater than," these postulates define the ordinal numerals. If  $a, b, c, \dots$ , are chicks, and  $>$  means "pecks," the postulates define the behavioral conditions under which a "pecking order" exists. If  $a, b, c, \dots$ , are minerals and  $>$  means "scratches," conformity to the postulates determines the existence of a unique scratching order. If it is suggested that aggression implies pecking or that hardness implies scratching, conformity of the behavior of chicks or minerals to the postulates indicates whether aggression or hardness, each assumed to be a single variable, can be measured on an ordinal scale in terms of observations of pecking or by scratching experiments.

Where the variable underlying the presumed order is discrete or where it is possible for two or more individuals to have identical or indistinguishably different amounts of this variable, we must enlarge the concept of order to include equality. To do so we define  $\succsim$  ("not greater than," "does not follow," "not older than," "does not scratch," "does not peck," "is not an ancestor of," etc.) and add the postulate:

If  $a \succsim b$  and  $b \succsim a$ , then  $a = b$ .

For example, two chicks are equally aggressive if neither one pecks the other, and two minerals are equally hard if neither scratches the other.

For an ordinal scale, the relation between the ordinal numerals and the attribute they measure is monotonic. If one individual has more of the attribute than another does, he must have a higher rank in any group of which they are both members. But *differences* in this attribute are not necessarily associated with proportionately equal differences in ordinal numerals. The measurements may in fact be replaced by their squares or their logarithms (if the measurements are positive), their cube roots, or any one of many other monotonic functions without altering their ordinal positions in the series.

Medians and percentiles of ordinal distributions are themselves seldom of interest because each of them merely designates an individual, or two individuals adjacent in the order. Hypotheses involving ordinal scale data may be tested by the Wilcoxon-Mann-Whitney, Kruskal-Wallis, Siegel-Tukey, and other procedures. (Interval-scale data are often

converted by ranking into ordinal scales to avoid the assumption of normality.) Both the Kendall and the Spearman rank correlation procedures apply to ordinal scales. [See NONPARAMETRIC STATISTICS, articles on ORDER STATISTICS and RANKING METHODS.]

*Quasi-ordinal scales.* Rankings made by judges are of course subject to errors. These are of two types: within-judge errors, where judges fail to discriminate consistently, and between-judge errors, where judges disagree in their rankings.

Within-judge discrimination can be evaluated, and the errors partially "averaged out," by the method of paired comparisons. A set of judgments such as  $a > b$ ,  $b > c$ ,  $c > a$ , which violates the third postulate, is termed a circular triad. If a judge's rankings provide some but not too many circular triads, a "best" single ranking is obtained by assigning a score of 1 to an individual each time he is judged better than another, summing the scores for each individual, and ranking the sums. This procedure yields a quasi-ordinal scale. A true ordinal scale exists only if there are no circular triads.

Between-judge agreement can be estimated by the coefficient of concordance or the average rank correlation, when several judges have ranked the same set of individuals on the same attribute. A true ordinal scale exists if all judges assign the same ranks to all individuals, so that the coefficient of concordance and the average rank correlation are both unity. If these coefficients are not unity but are still fairly high, the "best" single ranking of the individuals is obtained by summing all ranks assigned to each individual by the several judges, and then ranking these sums. This case is again a quasi-ordinal scale: an ordinal scale affected by some unreliability.

All statistical procedures which apply to ordinal scales apply also to quasi-ordinal scales, with the reservation that the results will be attenuated by scale unreliability.

*Interval scales.* An interval scale has equal units, but its zero point is arbitrary. Two interval scales measuring the same variable may use different units as well as different arbitrary zero points, but within each scale the units are equal. The classic examples are the Fahrenheit and centigrade scales of temperature used in physics.

For an interval scale, the relation between the scale numbers and the magnitudes of the attribute measured is not only monotonic; it is linear. Hence, if two interval scales measure the same variable in different units and from different zero points, the relation between them must be linear also. The

general linear equation, for variables  $X$  and  $Y$ , is of the form  $Y = a + bX$ . Thus, for the Fahrenheit and centigrade scales we have

$$^{\circ}\text{F} = 32 + 1.8^{\circ}\text{C}; \quad ^{\circ}\text{C} = .5556^{\circ}\text{F} - 17.7778.$$

On interval scales, differences in actual magnitudes are reflected by *proportional* differences in scale units. Thus, if two temperatures are 18 units apart on the Fahrenheit scale, they will be 10 units apart on the centigrade scale, no matter where they are located on these scales (e.g.,  $0^{\circ}$  to  $18^{\circ}\text{F}$  or  $200^{\circ}$  to  $218^{\circ}\text{F}$ ).

Interval-scale units may be added and subtracted, but they may not be multiplied or divided. We cannot say that Fahrenheit temperature  $64^{\circ}$  means twice as hot as Fahrenheit  $32^{\circ}$ . Almost all ordinary statistical procedures may be applied to interval-scale measurements, with the reservation that measures of central tendency must be interpreted as depending upon the arbitrary zero points. Almost all other statistical procedures are functions of deviations of the measures from their respective means and, hence, involve only addition and subtraction of the scale units. [See MULTIVARIATE ANALYSIS, articles on CORRELATION; STATISTICS, DESCRIPTIVE, article on LOCATION AND DISPERSION.]

*Quasi-interval scales.* Suppose that to each successive unit of an interval scale we apply a *random* stretch or compression. In this context randomness means that if the actual length of each unit is plotted against its ordinal position, there would be no trend of any sort: the larger units would not occur more frequently at one end, at both ends, in the middle, or in either or both intermediate regions. If the largest unit is small compared with the range of the variable in the group measured, we have a quasi-interval scale. All ordinary statistical procedures apply to quasi-interval scales, with the reservation that they have reduced reliability: errors of measurement are built into the scale units.

*Ratio scales.* A ratio scale has the properties of an interval scale and in addition a true zero point, the scale-value zero meaning absence of any amount of the variable measured. Classic examples from physics are length and weight. The relation between the actual quantities and the scale values is linear, and the equations, moreover, have no constants and are of the form  $Y = bX$ . If two ratio scales measure the same variable in different units, any measurement on one scale is the same multiple or fraction of the corresponding measurement on the other scale. Thus, if the length of an object is  $X$  inches, its length is also  $2.54X$  centimeters. And if the length of another object is  $Y$  centimeters, its length is also  $.3937Y$  inches.



Ratio-scale units may be multiplied and divided as well as added and subtracted. A man who is six feet high is exactly twice as tall as is one who is three feet high. Statistics applicable to ratio scales include geometric means, harmonic means, coefficients of variation, standard scores, and most of the common transformations (square root, logarithmic, arcsine, etc.) used to achieve improved approximations to normality of data distributions, homogeneity of variances, and independence of sampling distributions from unknown parameters. [See STATISTICAL ANALYSIS, SPECIAL PROBLEMS OF, article on TRANSFORMATIONS OF DATA.]

**Quasi-ratio scales.** A quasi-ratio scale is a ratio scale with random stretches and compressions applied to its units, in much the same way described for quasi-interval scales. All of the statistics appropriate to ratio scales apply also to quasi-ratio scales, with the reservation of reduced reliability.

**Operational definitions.** For many of the variables of the physical sciences, and some of the variables of the social sciences, the variable itself is defined by the operations used in measuring it or in constructing the measuring instrument. Thus length can be defined as what is measured by a ruler or yardstick. If we have first a "standard inch," say two scratches on a piece of metal (originally, according to story, the length of the first joint of a king's index finger), we can lay off successive inches on a stick by using a pair of dividers. For smaller units we can subdivide the inch by successive halving, using the compass and straight-edge of classical Euclidean geometry. Height, in turn, may be defined as length measured vertically, with vertical defined by a weight hanging motionless on a string.

Psychological and social variables, on the other hand, can less often be defined in such direct operational terms. For example, psychophysics and scaling have as major concerns the reduction of sensory, perceptual, or judgmental data to interval scale, quasi-interval scale, ratio scale, or quasi-ratio scale form. [See PSYCHOPHYSICS; SCALING.]

**Test scores as measurements.** In the context of this discussion, a test is usually simply a set of questions, often with alternative answers, printed in a paper booklet, together with the instructions given by an examiner; a test performance is whatever an examinee does in the test situation. The record of his test performance consists of his marks on the test booklet or on a separate answer sheet. If he is well motivated and has understood the examiner's instructions, we assume that the record of his test performance reflects his knowledge of

and ability in the field covered by the test questions. For the simpler types of items, with a simple scoring procedure, we credit him with the score +1 for each item correctly marked and 0 for each item incorrectly marked. We know that the organization of knowledge of an area in a human mind is complex and that an examinee's answer to one question is, in consequence, not independent of his answers to other questions. His score on the test is supposed to represent his total knowledge of the field represented by all the items (questions and alternative answers). In order to justify using such a score, we must be able to make at least two assumptions: (1) that knowledge of the area tested is in some sense cumulative along a linear dimension and (2) that there is at least a rough one-to-one relation between the amount of knowledge that each individual possesses and his test score.

The scores on a test, then, should form a quasi-ordinal scale. But suppose we have a 100-item, five-alternative, multiple-choice test with items arranged in order of difficulty from very easy to extremely hard. Individual A gets the first 50 right and a random one-fifth of the remainder. Individual B gets all the odd-numbered items right and a random one-fifth of all the even-numbered items. If a simple scoring formula that credits each correct response with a score of +1 is used, the score of each is 60. Yet most persons would tend to say that individual B has more of the ability measured by the test than has individual A. These cases are of course extreme, but in general we tend to attribute higher ability to an individual who gets more hard items right, even though he misses several easy items, than to one who gets very few hard items right but attains the same score by getting more easy items right. Thorndike (1926) distinguishes between *altitude of intellect* and *range of intellect* (at any given altitude) as two separate but correlated variables. He discusses ways of measuring them separately, but his suggestions have so far had little impact on the mainstream of psychometric practice. [Additional discussion of test scores as measurements is provided in the section "Item analysis," subsection on "Indexes of difficulty"; see also THORNDIKE.]

**Correction for guessing.** When an objective test is given to an individual, the immediate aim is to assess his knowledge of the field represented by the test items or, more generally, his ability to perform operations of the types specified by these items. But with true-false, multiple-choice, matching, and other *recognition*-type items, it is also possible for the examinee to mark right answers by

guessing. It is known that the guessing tendency is a variable on which large individual differences exist, and the logical purpose of the correction for guessing is to reduce or eliminate the expected advantage of the examinee who guesses blindly instead of omitting items about which he knows nothing. [See RESPONSE SETS.]

The earliest derivations were based on the *all-or-none assumption*, which holds that an examinee either knows the right answer to a given item with certainty and marks it correctly, or else he knows nothing whatever about it and the mark represents a blind guess, with probability  $1/a$  of being right (where  $a$  is the number of alternatives, only one of which is correct). Under this assumption we infer that when the examinee has marked  $a - 1$  items incorrectly, there were really  $a$  items whose answers he did not know, and that he guessed right on one of them and wrong on the other  $a - 1$ . Hence, for every  $a - 1$  wrong answers we deduct one from the number of right answers for the one he presumably got right by guessing. The correction formula is then

$$S = R - W/(a - 1),$$

where  $S$  is the corrected score (the number of items whose answers the examinee presumably knew),  $R$  is the right number, and  $W$  is the number wrong. It is assumed that this formula will be correct on the average, although in any particular case an examinee may guess the correct answers to more or fewer than  $W/(a - 1)$  items.

There is empirical evidence (Lord 1964) that correction for guessing corrects fairly well for high guessing tendency, but not so well for extreme caution, since the examinee is credited with zero knowledge for every item he omits. If an examinee omits items about which he has some, but not complete, knowledge, he will still be penalized. Hence, instructions should emphasize the point that an item should be omitted *only* if an answer would be a *pure guess*. If an examinee has a "hunch," he should always play it; and if he can eliminate one alternative, he should always guess among the remainder. This is a matter of ethics applying to all tests whose items have "right" answers; an examinee should never be able to increase his most probable score by disobeying the examiner's instructions.

A timed power test should begin with easy items and continue to items too hard for the best examinee. The time limit should be so generous that every examinee can continue to the point where his knowledge becomes substantially 0 for every re-

maining item. In this case the correction formula will cancel, on the average, the advantage that would otherwise accrue to those examinees who, near the end of the test period, give random responses to all remaining items.

There are only three conditions under which the correction for guessing need not be used: (1) there is no time limit, and examinees are instructed to mark every item; (2) the time limit is generous, examinees are instructed to mark every item as they come to it, and a minute or two before the end of the session the examiner *instructs* them to record random responses to all remaining items; (3) the test is a pure speed test, with no item having any appreciable difficulty for any examinee. In this case, errors occur only when an examinee works too fast.

### Item analysis

In the construction of standardized tests, item analysis consists of the set of procedures by which the items are pretested for difficulty and discrimination by giving them in an experimental edition to a group of examinees fairly representative of the target population for the test, computing an index of difficulty and an index of discrimination for each item, and retaining for the final test those items having the desired properties in greatest degree. *Difficulty* refers to how hard an item is, to how readily it can be answered correctly. A test item possesses *discrimination* to the extent that "superior" examinees give the right answer to it oftener than do "inferior" examinees. "Superior" and "inferior" are usually defined by total scores on the experimental edition itself. This is termed *internal consistency analysis*. When the less discriminating items are eliminated, the test becomes more *homogeneous*. A perfectly homogeneous test is one in which the function or combination of related functions measured by the whole test is also measured by every item. A test may, however, be designed for a specific use—e.g., to predict college freshman grades—in which case "superior" and "inferior" may be defined externally by the freshman grade-point average.

Wherever possible, the experimental edition is administered to the item-analysis sample without time limit and with instructions to the examinees to mark every item. If the experimental session has a time limit, the subset of examinees who mark the last few items form a biased subsample, and there is no satisfactory way to correct for this bias (Wesman 1949).

The experimental group for an item analysis



should be reasonably representative of the target population, and particularly representative with regard to age, school grade(s), sex, socioeconomic status, residence (city or country), and any other variables that might reasonably be expected to correlate substantially with total scores. Its range of ability should be as great as that of the target population. Beyond this, it does not have to be as precisely representative as does a norms sample. Item analyses based on a group from a single community (e.g., a city and the surrounding countryside) are often quite satisfactory if this community is representative of the target population on all of the associated variables.

There are two major experimental designs for item analysis. The first is called the upper-and-lower groups (ULG) design. On the basis of the total scores (or some external criterion scores) an upper group and a lower group are selected: usually the upper and lower 27 per cent of the whole sample, since this percentage is optimal. With the ULG design, the only information about the total scores (or the external criterion scores) that is used is the subgroup membership. Hence, this design calls for large experimental samples.

In the second design all the information in the data is used: for each item the distribution of total scores (or external criterion scores) of those who mark it correctly is compared with the distribution of those who mark it incorrectly. This is the item-total score (ITS) design. Here the sample size can be smaller.

**Indexes of difficulty.** With either design, a quasi-ordinal index of the difficulty of an item is provided by the per cent of the total sample who respond correctly. With the ULG design, a very slightly biased estimate is given by the average per cent correct in the upper and lower groups; a correction for this bias is found in the tables compiled by Fan (1952). For many purposes, however, an index of difficulty with units which in some sense form a quasi-interval scale is desired. With free-choice items, and the assumption that the underlying ability is normally distributed in the experimental sample, the normal deviate corresponding to the per cent correct yields a quasi-interval scale of difficulty. But under this assumption the distribution of difficulty of a recognition-type item will be skewed, with amount of skewness depending on the number of alternatives. The precise form of this distribution is not known. Common practice involves discarding items with difficulties not significantly higher than chance, even if they show high discrimination; redefining per cent correct as per cent correct above chance,

$p' = (p - 1/a)/(1 - 1/a)$ , where  $p$  and  $p'$  are now proportions rather than percentages; and treating as a very rough quasi-interval scale the normal deviates corresponding to these adjusted proportions.

Another method first replaces the raw total scores (or external criterion scores) with normalized standard scores to form a quasi-interval score scale. The distributions of these normalized standard scores for those who pass and fail the item are then formed and smoothed, and the difficulty of the item is taken as the score corresponding to the point of intersection of the two distributions. This is strictly an ITS procedure.

When item difficulties have a rectangular distribution ranging from easy for the least able examinee to hard for the most able, and when items are all equally discriminating on the average, the distribution of the test scores will be approximately the same as the distribution of the ability which underlies them; and these scores will form a quasi-interval scale. Almost the only tests which actually are so constructed are those for the measurement of general intelligence, such as the Stanford-Binet. Most tests have roughly normal, or at best mildly flat-topped, distributions of item difficulties. When applied to a group for which the mean item difficulty corresponds to the mean ability level and in which the ability is approximately normally distributed, the resulting score distribution tends to be flat-topped. Empirical data support this theoretical conclusion.

Tests constructed with all items of almost equal difficulty are useful for selection purposes; they have maximum reliability at the given ability level. With a rectangular distribution of item difficulties, a test is equally reliable at all scale levels, but its reliability at any one level is relatively low. With a normal or near-normal distribution of item difficulties, the reliability is at a maximum in the region of the modal difficulty and decreases toward the tails, but this decrease is less marked than it is in the case of a test whose items are all equally difficult.

Although scores on tests with near-normal distributions of item difficulties are frequently treated as forming quasi-interval scales, they should more properly be treated as forming only quasi-ordinal scales. All the strictures against treating percentile ranks as interval scales apply to such raw-score scales with only slightly diminished force.

**Indexes of discrimination in ULG design.** For some purposes we need only to eliminate items for which the number of right answers is not significantly greater in the upper group than in the lower

group, using the chi-square test of association. This procedure is often used in the selection of items for an achievement test.

In other cases we may wish, say, to select the 100 "best" items from an experimental test of 150. Here "best" implies a quasi-ordinal index of discrimination for each item. Widespread-tails tetrachoric correlations are often employed (Fan 1952; Flanagan 1939). The correlation indexes are statistically independent of the item difficulties. Where we may need quasi-interval scales, the Fisher  $z'$ -transformation is commonly applied to the widespread-tails tetrachoric correlation, yielding at least a crude approximation to an interval scale.

A less common procedure is to use the simple difference between the per cents correct in upper and lower groups as the index of discrimination. This index is precisely the percentage of cases in which the item will discriminate correctly between a member of the upper group and a member of the lower group (Findley 1956).

**Indexes of discrimination in ITS design.** With the ITS design, a  $t$ -test may be used to test the hypothesis that the mean total (or external criterion) scores of those who do and do not mark the right answer to the item are equal. If we cannot assume normality of the score distribution, we can replace the raw scores with their ranks and use the two-group Wilcoxon-Mann-Whitney test with only slight loss of efficiency. [See LINEAR HYPOTHESES, article on ANALYSIS OF VARIANCE, for a discussion of the  $t$ -test; see NONPARAMETRIC STATISTICS, article on RANKING METHODS, for a discussion of the Wilcoxon-Mann-Whitney test.]

To obtain an ordinal index of discrimination, the biserial, point-biserial, or Brogden biserial correlation (1949) between the item and the total (or external criterion) scores may be used. A crude approximation to interval scaling is given by applying the Fisher  $z'$ -transformation to the biserial or Brogden biserial correlations. [See MULTIVARIATE ANALYSIS, articles on CORRELATION.]

**Item analysis with wide-range groups.** Some tests are designed to be used over several consecutive ages or grades, and the mean growth of the underlying variable may be assumed to be roughly linear. In such cases an item may be very hard at the lowest level but very easy at the highest, and highly discriminating at one level but quite undiscriminating at another. In such cases we may plot for each item the per cent correct at successive ages or grades, or the intersections of the score distributions for those who pass and fail the item at each age or grade. Before using this latter procedure, the raw scores may be scaled by first assign-

ing normalized standard scores at each age or grade, assuming that the underlying variable is normally distributed at each level, and then combining them into a single scale by adjusting for the mean differences from age to age or grade to grade. An item is then retained if it shows a regular increase from age to age or grade to grade, or if it shows a large increase from any one age or grade to the next and no significant decrease at any other level. The scale difficulty of each item is the score level at which the per cent correct is 50, or for recognition items, the score level at which  $p'$ , defined as above, is .50.

**Two-criterion item analysis.** When a test is designed to predict a single external criterion, such as freshman grade-point average, success in a technical training course, or proficiency on a given job, we can do somewhat better than merely to select items on the basis of their correlations with the criterion measure. The original item pool for the experimental edition is deliberately made complex, in the hope that items of different types will assess different aspects of the criterion performance. The best subset of items will then be one in which the items have relatively high correlations with the criterion and relatively low correlations with the total score on the experimental edition. Methods of item selection based on this principle have been discussed by Gulliksen (1950, chapter 21) and by Horst (1936), using in both cases the ITS design.

**Inventory items.** Aptitude, interest, attitude, and personality inventories usually measure several distinct traits or include items of considerable diversity, subsets of which are scored to indicate similarity between the examinee's answer pattern and those of a number of defined groups. The items are usually single words or, more commonly, short statements, and the examinee marks them as applicable or inapplicable to him, true or false, liked or disliked, or statements with which he agrees or disagrees. The scoring may be dichotomous (like or dislike), trichotomous (Yes, ?, No), or on a scale of more than three points (agree strongly, agree moderately, uncertain, disagree moderately, disagree strongly). Often the statements are presented in pairs or triplets and the examinee indicates which he likes most and which least, or which is most applicable to him and which least applicable. The distinction between inventories based on internal analysis and those based on external criteria is a major one.

For internal analysis, the items are first allocated by judgment to preliminary subscales, often on the basis of some particular theoretical formulation. Each item is then correlated with every sub-



scale and reallocated to the subscale with which it correlates highest; items which have low correlations with all subscales are eliminated. If the subscales are theoretically determined, all items which do not correlate higher with the subscales to which they were assigned than to any other are eliminated. If the subscales are empirically determined, new subscale scores are computed after items are reallocated, and new item-subscale correlations are obtained; this process is repeated until the subscales "stabilize." Purely empirical subscales may also be constructed by rough factor analyses of the item data or by complete factor analyses of successive subsets of items. [See FACTOR ANALYSIS.]

For a *normative* scale, the job is finished at this point. (All aptitude and achievement tests form normative scales.) But when the statements are presented in pairs or triplets, they form an *ipsative* or partly *ipsative* scale. For a perfectly *ipsative* scale, the items of each subscale must be paired in equal numbers with the items of every other subscale, and only *differences* among subscale scores are interpretable. The California Test of Personality and the Guilford and Guilford-Martin inventories are examples of normative scales. The Edwards Personal Preference Schedule is perfectly *ipsative*, and the Kuder Preference Record is partly *ipsative*.

In filling out inventories of these types, whose items do not have right or wrong answers, we want examinees to be honest and accurate. Normative inventories are easily fakeable; *ipsative* inventories somewhat less so. Response sets also affect inventory scores much more than they do aptitude and achievement test scores. Most of the better inventories therefore have special scales to detect faking and to correct for various types of response sets. In forming pairs and triplets, efforts are made to equalize within-set social desirability or general popularity, while each statement represents a different subscale. [See ACHIEVEMENT TESTING; APTITUDE TESTING; PERSONALITY MEASUREMENT, *articles on* PERSONALITY INVENTORIES and THE MINNESOTA MULTIPHASIC PERSONALITY INVENTORY; RESPONSE SETS; VOCATIONAL INTEREST TESTING.]

Inventories constructed on the basis of external criteria use a base group, usually large ("normal" individuals, "normal" individuals of given sex, professional men in general, high school seniors in general, high school seniors of given sex, and the like), and a number of special groups (hospital or clinic patients with the same clear diagnosis, or men or women in a given occupation). An answer (alternative) is scored for a special group if the

people in that group mark it significantly more often (scored positively) or significantly less often (scored negatively) than do the people in the base group. In some inventories the more highly significant or highly correlated answers are given greater positive or negative weights than are the less significant or less highly correlated answers. Inventories of this type are almost always normative, and new subscales can be developed whenever new special groups can be identified and tested. The outstanding examples are the Strong Vocational Interest Blank and the Minnesota Multiphasic Personality Inventory.

In inventories of this type, the same item may be scored for several subscales. In consequence there are inter-key correlations, and the reliabilities of differences between pairs of subscale scores vary with the numbers of common items. A further consequence is that general subscales based on factor analyses of individual subscale intercorrelations are difficult to evaluate, since the individual subscale scores are not experimentally independent. Similar difficulties arise in the interpretation of factor analyses of *ipsative* and partly *ipsative* scale scores.

### Reliability

Reliability is commonly defined as the accuracy with which a test measures whatever it does measure. In terms of the previous discussion, it might be defined in some cases as a measure of how closely a quasi-ordinal or quasi-interval scale, based on summation of item scores, approximates a true ordinal or interval scale. The following treatment assumes quasi-interval scales, since reliability theories based entirely on the allowable operations of ordinal arithmetic, which do not define the concepts of variance and standard deviation, have not been worked out. However, definitions and results based on correlations probably apply to quasi-ordinal scales if the correlations are Spearman rank correlations. [See MULTIVARIATE ANALYSIS, *articles on* CORRELATION.]

The raw score of an individual on a test may be thought of as consisting of the sum of two components: a true score representing his real ability or achievement or interest level or trait level, and an error of measurement. Errors of measurement are of two major types. One type reflects the limitation of a test having a finite number of items. Using ability and aptitude tests as the basis for discussion, the individual's true score would be his score on a test consisting of *all* items "such as" the items of the given test. On the finite test he may just happen to know the right answers to a greater

or lesser proportion of the items than the proportion representing his true score. Errors of this type are termed *inconsistency errors*. A second type of error reflects the fact that the *working* ability of an individual fluctuates about his *true* ability. On some occasions he can "outdo himself": his working ability exceeds his true ability. On other occasions he cannot "do justice to himself": his working ability is below his true ability. Working ability fluctuates about true ability as a result of variations in such things as motivation, inhibitory processes, physical well-being, and external events that are reflected in variation in concentration, cogency of reasoning, access to memory, and the like. Such fluctuations occur in irregular cycles lasting from a second or two to several months. Errors of this type are termed *instability errors*.

If a second test samples the same universe of items as does the first and in the same manner (random sampling or stratified random sampling with the same strata), the two tests are termed *parallel forms* of the same test. Parallel forms measure the same true ability, but with different inconsistency errors.

The basic theorem which underlies all formulas of reliability, and of empirical validity as well, may be stated as follows: *In a population of individuals, the errors of measurement in different tests and in different forms of the same test are uncorrelated with one another and are uncorrelated with the true scores on all tests and forms.*

**Coefficient and index of reliability.** The *reliability coefficient*,  $R$ , may be defined as the ratio of the true score variance to the raw score variance; it is also the square of the correlation between the raw scores and the true scores. The *index of reliability* is the square root of the reliability coefficient; it is the ratio of the standard deviation of the true scores to the standard deviation of the raw scores, or the correlation between the true scores and the raw scores. These definitions are purely conceptual. They are of no computational value because the true scores cannot be measured directly.

Furthermore, where  $R_A$  and  $R_B$  are the reliability coefficients of the two parallel forms and  $\rho_{AB}$  is the correlation between them, it is implied in the basic theorem that  $\sqrt{R_A R_B} = \rho_{AB}$ . If, moreover, as is usually the case, the two forms are equally reliable,  $R_A = R_B = \rho_{AB}$ ; i.e., the correlation between the two forms is the reliability coefficient of *each* of them. When we estimate  $\rho_{AB}$  by computing  $r_{AB}$ , the correlation in a sample, the estimate is not unbiased, but the bias is usually small if the sample is reasonably large.

**Consistency coefficient.** If two equally reliable parallel forms of a test are administered *simultaneously* (e.g., by merging them as odd and even items in the same test booklet), the reliability coefficient becomes a *consistency coefficient*, since instability errors affect both forms equally.

The split-half correlation (e.g., the correlation between odd and even items) provides the consistency coefficient of each of the half-tests. The consistency coefficient of the whole test, as estimated from the sample, is then derived from the Spearman-Brown formula:

$$R_{A+B} = \frac{2r_{AB}}{1 + r_{AB}},$$

where  $r_{AB}$  is the correlation between the half-tests. The more generalized Spearman-Brown formula is

$$R_n = \frac{n r_{AB}}{1 + (n-1) r_{AB}},$$

where  $r_{AB}$  is again the correlation between the half-tests and  $R_n$  is the consistency coefficient of a parallel form  $n$  times as long as one half-test. In deriving the Spearman-Brown formula, we must assume that the half-tests are equally variable as well as equally reliable, but these requirements are not very stringent. Kuder and Richardson (1937) also present several formulas for the consistency of one form of a homogeneous test. Their most important formula (formula 20) was generalized and discussed at some length by Cronbach (1951).

**Interform reliability coefficient.** If two equally reliable parallel forms of a test are administered to the same group of examinees at two different times, the correlation between them is an *interform reliability coefficient*. The interform reliability is lower than the consistency because it is affected by instability errors, which increase with time. Reports of interform reliability should include the length of the time interval.

**Stability coefficient.** Instability errors are related to the interval between testings and are independent of the inconsistency errors. The stability coefficient may be defined as the interform reliability that would be found if both forms of the test were perfectly consistent. It may be estimated by the formula

$$s_{AB} = r_{AB} / \sqrt{C_A C_B},$$

where  $r_{AB}$  is the interform reliability coefficient and  $C_A$  and  $C_B$  are the consistency coefficients of the two forms, each computed from the split-half correlation and the Spearman-Brown formula, or from the Kuder-Richardson formula 20. Its value is independent of the lengths of the two forms of



the test but dependent upon the time interval separating their administration.

The increase in interform reliability resulting from increase in test length may be estimated by the formula

$$R_n = \frac{n r_{AB}}{1 + (n-1) C_{AB}},$$

where  $r_{AB}$  is the interform reliability and  $C_{AB}$  is the consistency of each form. The two forms must be assumed equally consistent, and  $C_{AB}$  is computed as  $C_{AB} = \sqrt{C_A C_B}$ . Then  $R_n$  is an estimate of the interform reliability of a parallel form  $n$  times as long as one of the two actual forms. If  $n = 2$ , the formula

$$R_{A+B} = \frac{2r_{AB}}{1 + C_{AB}}$$

gives the interform reliability of scores on the two forms combined.

**Test-retest correlation.** When the same form is given to the same examinees on two different occasions, the correlation is *not* a stability coefficient, and it would not be a stability coefficient even if every examinee had total amnesia of the first testing (and of nothing else) on the second occasion. In addition to the quantitative fluctuations in working ability which give rise to instability errors, there are qualitative fluctuations in perceptual organization, access to memory, and reasoning-procedure patterns. In consequence, the same set of items, administered on different occasions, gives rise to different reactions; and in consequence there are still some inconsistency errors. Perseveration effects, including but not limited to memory on the second occasion of some of the responses made to particular items on the first occasion, introduce artificial consistency, in varying amounts for different examinees. In consequence, test-retest coefficients cannot be clearly interpreted in terms of reliability theory.

**Standard error of measurement.** If several parallel forms, all equally reliable and with identical distributions of item difficulty and discrimination, could be given simultaneously to one examinee, the standard deviation of his scores would be the standard error of measurement of one form for him. Such a standard error of measurement is an estimate of the *average* standard error of measurement for all members of an examinee group. The formula is

$$SE_m = s \sqrt{1 - r_{AB}},$$

where  $s$  is the standard deviation of the total scores and  $r_{AB}$  is their consistency, computed by the split-half correlation and the Spearman-Brown formula.

The standard error of measurement may also be defined, for the whole sample or population of examinees, as the standard deviation of the inconsistency errors or the standard deviation of the differences between raw scores and the corresponding true scores. With this last definition we can also compute a standard error of measurement which includes instability errors over a given time period, by letting  $s$  represent a pooled estimate of the standard deviation of one form based on the data for both forms, and  $r_{AB}$  an interform reliability coefficient. In this case,  $SE_m$  is the standard error of measurement of one (either) form.

**Reliability and variability.** The variance of true scores increases with the variability of ability in the group tested, while the variance of the errors of measurement remains constant or almost constant. By the variance-ratio definition of the reliability coefficient, it follows that this coefficient increases as the range of ability of the group measured increases. The reliability coefficient of a particular test is higher for a random sample of all children of a given age than for a sample of all children in a given grade, and lower for all children in a single class. When a reliability coefficient is reported, therefore, the sample on which it was computed should be described in terms which indicate as clearly as possible the range of ability of the subjects.

The formula relating variability to reliability is

$$R_{AB} = 1 - \frac{s^2}{S^2} (1 - r_{AB}),$$

where  $s^2$  and  $r_{AB}$  are the variance and the reliability coefficient of the test for one group, and  $S^2$  and  $R_{AB}$  are the variance and reliability coefficient for another group. The group means should be similar enough to warrant the assumption that the average standard error of measurement is the same in both groups. If a test author reports  $r_{AB}$  and  $s^2$  in his manual, a user of the test need only compute  $S^2$  for his group, and  $R_{AB}$  for that group can then be computed from the formula. Note that this formula applies exactly only if the test-score units form a quasi-interval scale over the whole score range of both groups.

**Reliability at different score levels.** If the test-score units do not form a quasi-interval scale, the standard error of measurement will be different at different score levels. If two forms of the test or two halves of one form are *equivalent*, and the experimental sample is large enough, the standard error of measurement may be computed for any given score level. Two parallel forms or half-tests are equivalent if their joint distributions of item

difficulty and item discrimination are essentially identical. In this case, their score distributions will also be essentially identical.

To compute the standard error of measurement at a given score level, we select from a large experimental sample a subgroup whose total scores on the two forms or half-tests combined are equal within fairly narrow limits. The standard error of measurement of the total scores is then the standard deviation of the differences between the half-scores. When, as is usually the case, the half-scores are based on splitting one form into equivalent halves administered simultaneously, the standard errors of measurement at different score levels are based only on inconsistency errors.

The reliability coefficient at a given score level, still referred to the variability of the whole group, is given by

$$r_{AB} = 1 - SE_m^2/s^2,$$

where  $SE_m$  is defined as above and  $s^2$  is the total-score variance of the whole group.

**Comparability.** Two forms of a test, or two tests measuring different true-score functions, are termed comparable if and only if their units of measurement are equal. If the units do not form quasi-interval scales, they can be made comparable only if their score distributions are of the same shape and their standard errors of measurement are proportional at all score levels. Only equivalent forms have comparable raw-score units.

If two different tests have proportionally similar joint distributions of item difficulties and discriminations, they will meet these conditions. Meaningful interpretations of profiles of scores on different tests can be made only if the scores are comparable.

### Validity

Test validity has to do with *what* a test measures and how well it measures what it is intended to measure, or what it is used to measure in any particular application if it is a multiple-use test.

**Content validity.** Content validity applies mainly to achievement tests, where the questions themselves define the function or combination of related functions measured and there is no external criterion of achievement with which the scores can be compared. The test developer should provide a detailed outline of both the topics and the processes—such as information, comprehension, application, analysis, synthesis, evaluation, etc.—that the test measures. A more detailed list of processes, with illustrative test questions from several fields, is given in *Taxonomy of Educational Objectives* (1963–1964). The item numbers of the test items

are then entered in the cells of the outline, along with their indexes of difficulty and discrimination.

Evaluations of content validity are essentially subjective. The prospective user of the test may agree or disagree to a greater or lesser extent with the outline or the basis on which it was constructed, with the allocation of items to topics and processes, or with the author's classification of some of the items. If all such evaluations are positive, the test's validity is equal, for all practical purposes, to its interform reliability over some reasonable time period.

In constructing an achievement test, item analysis ordinarily consists only of the elimination of nondiscriminating items. If the test is to yield a single score, the various topics and processes must be sufficiently homogeneous to permit every item to correlate positively and significantly with the total score. All further elimination occurs in the balancing of the item contents against the requirements of the topic-by-process outline. In discussing school achievement tests, content validity is often termed curricular validity.

**Empirical validity.** Empirical validity is concerned with how well a test, either alone or in combination with others, measures what it is used to measure in some particular application. The empirical validity of a test is evaluated in terms of its correlation with an external *criterion* measure: an experimentally independent assessment of the trait or trait complex to be predicted. The term "prediction" is used here, without regard to time, to designate any estimate made from a regression equation, expectancy table, or successive-hurdles procedure. The term "forecast" will be used when we explicitly predict a *future* criterion. Empirical validity is also termed statistical validity and criterion validity.

There are two basically different types of criteria. The first may be termed *sui generis* criteria, criteria that exist without any special effort made to predict them. Examples include persistence in college, success or failure in a training course, dollar volume of sales, years of service in a company, and salary level. The unreliability of the criterion measure sets a natural upper limit for the validity of any predictor. The validity of a predictor or predictor battery is simply its correlation or multiple correlation with the criterion. We term such a correlation an index of *raw validity*.

The second type of criteria may be termed *constructed criteria*, and are developed upon the basis of a trait concept such as academic ability, job proficiency, or sales accomplishment. For academic achievement such a criterion might be grade-point



average in academic subjects only. For job proficiency it might be based on quantity of output, quality of output, material spoilage, and an estimate of cooperation with other workers. For sales accomplishment it might be based on number of sales, dollar volume, new customers added, and an estimate of the difficulty of the territory. In any event, it must be accepted as essentially an *operational definition* of the trait concept and, hence, intrinsically content-valid. And since the error of measurement is no part of the operational definition of a trait concept, it is evident that we should predict *true* criterion scores rather than *raw* scores.

*Concurrent and forecast validity.* We can recognize two types of assessments of true validity: *concurrent* true validity and *forecast* true validity. In the first case, the criterion measure is usually one which is expensive or difficult to obtain, and the predictor is designed to be a *substitute measure* for it. In this case the predictor test or battery should be administered at the *middle* of the time interval over which the criterion behavior is observed; otherwise, instability errors will distort validity estimates.

For *forecast* true validity, the predictor test or battery is administered at some "natural" time: at or shortly before college entrance or admission to training or initial employment, and the criterion data should cover the later time period over which they will be most valid in terms of the criterion trait concept. In this case, the instability errors resulting from the earlier administration of the predictor test or battery are intrinsic to the prediction enterprise.

To obtain a quick rough estimate of forecast value, an investigator often tests present employees, students, or trainees at essentially the same time that the criterion data are obtained. This procedure should *not* be termed concurrent validation but, rather, something like *retroactive* validation.

*Test selection and cross-validation.* A common type of empirical validation study consists in administering to the experimental sample more predictors than are to be retained for the final battery. The latter then consists of that subset of predictors, of manageable length, whose multiple correlation most nearly approximates that of the whole experimental battery. Predictors commonly include scored biographical inventories, reference rating scales, and interview rating scales as well as tests.

When a subset of predictors is selected by regression procedures, its multiple correlation with the criterion is inflated: sampling error determines in part which predictors will be selected and what their weights will be. In *cross-validation*, the re-

duced predictor battery is applied to a second criterion group, using the weights developed from the first group. The aggregate correlation in the second group (now no longer a multiple correlation) is an unbiased estimate of the battery validity. In estimating forecast true validity, two criterion measures for each examinee are required only in the cross-validation sample. [See LINEAR HYPOTHESES, article on REGRESSION.]

The same situation arises in even more exaggerated form when predictor items are selected on the basis of their correlations with an external criterion measure. Each predictor requires a separate item-analysis sample. A different sample is required to determine predictor weights. And a still different sample is required to estimate the validity of the whole predictor battery.

Various split-sample methods have been devised to use limited data more effectively. Thus, test selection may be carried out on two parallel samples, keeping finally the subset of tests selected by both samples. The validity of the battery is estimated in each sample by using the weights from the other. The average of the two validity indexes is then a lower bound for the battery validity when the weights used are the averages of the weights from the two samples.

*Validation procedures.* The commonest methods of test selection and use are those described above, using multiple regression as the basic procedure. These procedures assume that the criterion elements are all positively and substantially correlated, and can be combined with suitable weights into a single criterion measure of at least moderate homogeneity. It is then further assumed that a low score on one predictor can be compensated for by high scores on others, a weighted total score on all predictors being accepted as a single predictor of criterion performance.

Some criteria, however, consist of elements which are virtually uncorrelated. In such cases the elements must be predicted separately. In practice there is usually one predictor for each element, although it is possible to predict an element by its multiple regression on two or more predictors. In this situation, the preferred procedure is the multiple cutoff procedure. Each criterion-element measure is dichotomized at a critical (pass-fail) level, the corresponding predictor level is determined, and a successful applicant must be above the critical levels on *all* predictors. A further refinement consists in rating the criterion elements on their importance to the total job and requiring a successful applicant to be above the critical levels on the predictors of all the more important elements, but

permitting him to be a little (but not far) below this level on one or two of the predictors of the less important elements.

In predicting a dichotomous criterion, the most accurate predictions are made when the predictor cutoff score is at the point of intersection of the smoothed frequency curves of predictor scores for the upper and lower criterion groups. If the applicant group is large, however, the predictor cutoff score may be set one or two standard errors of measurement above this point.

*Correction for attenuation.* Correction for attenuation is a procedure for estimating what the correlation between two variables would be if both of them were perfectly reliable or consistent; i.e., if the correlation were based on the true scores of both variables. The unreliabilities of the variables attenuate (reduce) this correlation. To determine the proper correction, the experimental design must be such that the instability errors in the intercorrelation(s) are identical with those in the reliability or consistency coefficients.

*Index and coefficient.* In discussing formulas for validity, the term "index" rather than the term "coefficient" has been used, although the latter is the term commonly used. The square of each of these correlations, however, is a *coefficient of determination* (of raw criterion scores by raw predictor scores, true criterion scores by raw predictor scores, or true criterion scores by true predictor scores). The reliability coefficient is also a coefficient of determination (of true scores by raw scores on the same variable), and the *index* of reliability is its square root.

As the intrinsic validity (the true validity of a perfectly reliable predictor) approaches unity, the index of true validity approaches the index of reliability of the predictor, since in this case the predictor and criterion true scores are identical; and an unreliable predictor cannot predict anything better than it predicts its own true scores. The statement "The upper limit of a test's validity is the square root of its reliability" is erroneous: the upper limit of its validity is its reliability when both are expressed either as indexes or as coefficients. The error is due to the common practice of calling "indexes" of validity "coefficients" of validity.

*Synthetic validity.* Synthetic validation is test selection without a criterion variable; and synthetic validity is an estimate of empirical validity, also without a criterion variable. The procedure is based on job analysis and the accumulated experience from many previous empirical validation studies. The number of possible jobs, and the number of

real jobs as well, greatly exceeds the number of distinct job elements and job-qualification traits. If previous studies have shown which qualification traits are required for each job element of a given job, and what predictors best predict each of these traits, predictors can be selected for the given job on the basis of a job analysis without a new empirical study, and rough estimates can even be made of the probable validity of the prediction battery. The procedures of job analysis are by now fairly well refined; there are substantial bodies of data on the qualification requirements for many job elements; and there are also fairly large amounts of data on the correlations between predictors and qualification traits and the intercorrelations among such predictors. Hence, synthetic validation is by now at least a practical art, and its methodology is approaching the status of an applied science. Synthetic validation is the only procedure which *can* be used when the number of positions in a given job category is too small to furnish an adequate sample for an empirical validation study.

*Factor analysis of validity.* Factor analysis provides a way of answering the question "What does this test measure?" The simplest answer is merely a collection of correlations between the given test and a variety of other better-known tests. However, if the given test and several others have all been administered to the same large sample, the factor structure of the given test provides a much better answer.

When a large number of tests are administered serially in a factor analysis study, the instability errors are greater for tests far apart in the series than for tests administered consecutively. This leads to the generation of small instability factors and complicates making the decision about when to stop factoring. If there are two forms of each test, all the A forms might be given serially, and then after an interval of a week or more all the B forms might be given in a different serial order. Two parallel factor analyses could then be performed, one (for tests 1 and 2, say) using the correlation between 1A and 2B; the other, the correlation between 1B and 2A. The correlations between 1A and 2A and between 1B and 2B would not be used. Interform reliabilities consistent with the intercorrelations would be given by the correlations between 1A and 1B and between 2A and 2B. [See FACTOR ANALYSIS.]

*Construct validity.* Construct validation is an attempt to answer the question "Does this test measure the trait it was designed to measure?" when no single criterion measure or combination of criterion measures can be considered a well-defined,



agreed-upon, valid measure of the trait; and there is in fact no assurance that the postulated trait is sufficiently unitary to be measurable. We start with a trait *construct*: a hypothesis that a certain trait exists and is measurable. Then we build a test we hope will measure it. There are two questions, to be answered simultaneously: (1) Does the trait construct actually represent a measurable trait? (2) Does the test measure that trait rather than some other trait?

From the trait construct we draw conclusions about how the test should correlate with other variables. With some it ought to correlate fairly highly, and if in every such case it does, the resulting evidence is termed *convergent* validity. Variables of this type are in a sense *partial* criteria. With other variables it should have low correlations, and if in every such case it does, the resulting evidence is termed *discriminant* validity. Consider the trait construct "general intelligence" and the Stanford-Binet Scale.

(a) General intelligence should increase with age during the period of growth. Mean scores on the Stanford-Binet increase regularly throughout this period, but so do height and weight.

(b) General intelligence can be observed and rated with some validity by teachers. Children rated bright by their teachers make higher Stanford-Binet scores than do children of the same age rated dull. There is some judgmental invalidity, however; docile and cooperative children are overrated, and classroom nuisances are underrated.

(c) The extremes of general intelligence are more certainly identifiable. Individuals judged mentally deficient make very low scores, but so do prepsychotics and children with intense negative attitudes toward school. Outstanding scholars, scientists, writers, and musical composers make high scores. Equally outstanding statesmen, executives, military leaders, and performing artists make somewhat lower but still quite high scores. This finding also agrees with the hypothesis, for people in the latter categories need high, but not so *very* high, general intelligence; and they also need special talents not so highly related to general intelligence.

(d) Items measuring diverse cognitive traits should correlate fairly highly with one another and should generate a large general factor if general intelligence is indeed a relatively unitary measurable trait. The items of the Stanford-Binet clearly do measure diverse cognitive traits, and their intercorrelations do generate a large general factor.

(e) Reliable homogeneous tests of clearly cognitive traits should correlate fairly highly with general intelligence. Tests of vocabulary, verbal and nonverbal reasoning, arithmetic problems, and

visual-space manipulation correlate fairly highly with the Stanford-Binet, and not quite so highly with one another.

(f) Tests judged "borderline cognitive" should have positive but moderate correlations with general intelligence. Tests of rote memory, verbal fluency, mechanical knowledge, and the like do have positive but moderate correlations with the Stanford-Binet.

(g) Wholly noncognitive tests should have near-zero correlations with general intelligence. Tests of writing speed, visual acuity, physical strength, and the like do have near-zero correlations with the Stanford-Binet.

The full combination of these predictions and results, along with others not cited above, leads us to place considerable confidence in the trait status of the construct "general intelligence" and in the Stanford-Binet Scale (along with others) as a measure of this trait.

If there is even a single glaring discrepancy between theory and data, either the theory (the trait construct) must be revised, or the test must be considered invalid. A test of "social intelligence" was shown to correlate as highly with a test of general-verbal intelligence as other tests of general-verbal intelligence did. From this one finding we must conclude either that social intelligence is not a trait distinct from general-verbal intelligence, or that the test in question is not a discriminatingly valid measure of social intelligence because it is a valid measure of general-verbal intelligence. [See INTELLIGENCE AND INTELLIGENCE TESTING.]

Construct validity is, in the end, a matter of judgment and confidence. The greater the quantity of supporting evidence, the greater the confidence we can place in both the trait construct and the test. But the possibility of a single glaring discrepancy is never wholly ruled out by any quantity of such evidence (see especially Cronbach & Meehl 1955).

### Test norms

Since the raw scores on educational and psychological tests consist of arbitrary units with arbitrary zero points—units which in most cases vary systematically as well as randomly with score level—these individual scores can be interpreted intelligently only by comparing them with the *distributions* of scores of defined groups, or norms. The comparisons are facilitated by using various types of score transformations based on these distributions.

Norms may be local, regional, or national; they may refer to the whole population or to defined subgroups, such as sex and occupation. Local norms

are usually determined by testing everyone in a particular group: all children in certain grades of a school system, all freshmen applying to or admitted to a college or university, all employees in specified jobs in a company, etc. Regional and national norms must be determined by sampling. Since random sampling of the whole defined population is never practical, much care is necessary in the design of the sampling procedure and in the statistical analysis of the data to assure the representativeness of the results. The principles and procedures of survey sampling are beyond the scope of this article, but one caution should be noted. Use of a "pick-up" sample, depending upon the vagaries of cooperation in various communities, however widespread over regions, rural and urban communities, etc., followed by weighting based on census data, is never wholly satisfactory, although it is the method commonly employed. When norms are based on a sample which omits some major regions or population subgroups entirely, no weighting system can correct the bias, and such norms must be used with extreme caution in these other regions and with omitted subgroups. [See SAMPLE SURVEYS.]

**Grade norms.** For elementary and junior high school achievement tests, grade norms are commonly employed. The data used are the mean or median scores of children in successive grades, and the unit is one-tenth of the difference between these averages for successive grades. There is probably some error in assuming one month of educational growth over the summer vacation, but this error in interpreting grade scores is small in comparison with the standard errors of measurement of even the best achievement tests.

**Age norms.** Age scores are used mainly with general intelligence tests, where they are termed mental ages. The data are the average scores of children within one or two months (plus or minus) of a birthday, or the average scores of all children between two given birthdays, and the unit is one-twelfth of the difference between the averages for successive ages; i.e., one month of mental age. For ages above 12 or 13, extrapolations and corrections of various sorts are made, so that mental ages above 12 or 13 no longer represent age averages but, rather, estimates of what these averages would be if mental growth did not slow down during the period from early adolescence to maturity. Thus, when we say that the mental age of the average adult is 15, we do *not* mean that for most people mental growth ceases at age 15. What we do mean is that the total mental growth from age 12 to maturity is about half the mental growth from age 6 to age 12; so that at age 15 the average per-

son would have reached the mental level he actually reached at maturity (in the early or middle twenties) if mental growth from 12 to 15 had continued at the same average rate as from 6 to 12.

Age norms have been used in the past with elementary and junior high school achievement tests also, defining "arithmetic age," "reading age," "educational age," etc., in like manner; but they and the corresponding "arithmetic quotient," "reading quotient," and "educational quotient" are no longer used.

**Quotient scores.** The IQ as defined originally, for children up to age 12 or 13, is  $100(MA/CA)$ , the ratio of mental age to chronological age, multiplied by 100 to rid it of decimals. For older ages the divisor is modified so that the "equivalent chronological age" goes from 12 or 13 to 15 as the actual age goes from 12 or 13 to mental maturity. The *de facto* corrections have usually been crude.

The IQ would have the same standard deviation at all ages if the mental growth curve were a straight line, up to age 12 or 13, passing through true zero at birth, and if the standard deviation of mental ages were a constant fraction of the mean mental age at all chronological ages. These conditions are met very roughly by age scales such as the Stanford-Binet and are not met at all well by any other tests. In consequence, and because of the troubles encountered in the extrapolation of mental age scales and the derivation of equivalent chronological ages beyond age 12 or 13, IQs are no longer computed for most tests by the formula  $100(MA/CA)$ . They have been replaced quite generally by "deviation IQs," which are standardized scores or normal-standardized scores for age groups; and even the names "intelligence" and "IQ" are tending to be replaced by other names because of their past widespread misinterpretation as measuring innate intellectual capacity. [See INTELLECTUAL DEVELOPMENT; INTELLIGENCE AND INTELLIGENCE TESTING.]

**Modal-age grade scores.** As we proceed from the first grade to the ninth, the age range within a grade increases, and the distribution of ages becomes skewed, with the longer tail in the direction of increased age. The reason is that retardation is considerably more common than acceleration, and the consequence is that in the upper grades the grade averages on tests are no longer equal to the averages for pupils making normal progress (one grade per year) through school. Modal-age grade scores are used to compare the level of an individual child's performance with the level representing normal school progress. The data are the means or medians, not of all children in a given grade but, rather, of those children whose ages are within



six to nine months of the *modal* age (not the median or mean age) for the grade. The units are otherwise the same as those of total-group grade scores, with all the interpretive difficulties noted previously. Modal-age grade scores are recommended for judging the progress of an individual child; total-group grade scores for comparing classes, schools, and school systems.

When total-group grade scores are based on grade medians, they are about one-third closer to the corresponding modal-age grade scores than when they are based on grade means.

**Standard scores and standardized scores.** When individual scores are to be compared with those of a single distribution, rather than with the means or medians of successive groups (as is the case with grade and age scores), scores based on the mean and standard deviation of the distribution are frequently employed. Originally, standard scores were defined by the formula  $Z = (x - \bar{x})/s$ , where  $x$  is a raw score,  $\bar{x}$  is the group mean, and  $s$  is the standard deviation. Thus Z-scores have a mean of zero and a standard deviation of unity; the variance is also unity, and the product-moment correlation between two sets of Z-scores is the same as the covariance:  $r_{AB} = \sum Z_A Z_B / N$ . For raw scores below the mean, Z-scores are negative. Thus the Z-score 1.2 corresponds to the raw score which is 1.2 standard deviations above the mean, and the Z-score  $-.6$  corresponds to the raw score which is six-tenths of a standard deviation below the mean.

Because of the inconveniences in using negative scores and decimals, Z-scores are usually converted via a linear transformation into some other system having an arbitrary mean and an arbitrary standard deviation. These other systems are commonly termed standard-score systems, but the present writer prefers, like Ghiselli (1964), to reserve the term "standard score" for Z-scores and to call the other systems "standardized scores."

The units of a standardized-score system form a quasi-interval scale if and only if the raw scores form such a scale. When the item-difficulty distribution is roughly normal rather than rectangular, the units are smallest near the score corresponding to the modal difficulty, and they become progressively larger with distance in either direction from this level.

**Normal-standardized scores.** Normal-standardized scores are standard or standardized scores for a normal distribution having the same mean and standard deviation as the raw-score distribution. They are found by looking up in a table of the normal distribution the Z-scores corresponding to the percentile ranks of the raw scores in the actual distribution, and then subjecting these Z-scores to

any desired arbitrary linear transformation. This procedure corrects for departure of the score distribution from normality, but it does not insure equality of units in any practical sense, even if the distribution of the underlying ability is also normal.

The phrase "normal-standardized scores" is to be preferred to the more common "normalized standard scores." To a mathematician, "standardizing" means reducing to Z-scores, and "normalizing" means producing scores each equal to  $Z/\sqrt{N}$ , with sum of squares (instead of standard deviation) unity, and has no reference to the normal distribution.

**Percentiles and percentile ranks.** A percentile is defined as that score (usually fractional) below which lies a given percentage of a sample or population. The median is the 50th percentile: half the group make scores lower than the median and half make scores higher. The lower quartile is the 25th percentile and the upper quartile is the 75th percentile. All score distributions are grouped distributions, even though the grouping interval may be only one score unit. Percentiles are computed by interpolation under the assumption that the abilities represented by the scores within an interval are evenly distributed across that interval.

Percentile ranks are the percentiles corresponding to given scores. Since a single score represents an *interval* on a continuous scale, its percentile rank should be the percentage of individuals who make lower scores plus half the percentage who make the given score. In practice they are frequently computed as simply the percentage who make lower scores, and occasionally as the percentage who make the same or lower scores. Neither of these errors is large in comparison with the error of measurement.

Percentiles and percentile ranks are sometimes given for grade groups, age groups, or normal-age grade groups with elementary and junior high school achievement tests and intelligence tests. They are used more commonly with high school and college tests, with the reference group all students in a given class (grade), or all college applicants, in the case of college entrance tests. For tests in particular subject areas, the reference groups are more commonly all students who have studied the subject for a given number of years in high school or college.

Strict warnings are commonly given against treating percentile ranks as though they form quasi-interval scales; but as noted above, raw scores, standardized scores, and normal-standardized scores may be little, if any, better in this respect when item-difficulty distributions are far from rectangu-

lar. It is quite possible, in fact, that for some not uncommon item-difficulty distributions, the percentile ranks may have more nearly the properties of an interval scale than the raw scores have.

**Score regions.** Centiles and deciles are the regions between adjacent percentiles and sets of ten percentiles. The first centile is the score region below the first percentile. The 100th centile is the region above the 99th percentile. The  $k$ th centile is the region between the  $(k-1)$ th and  $k$ th percentiles. The first decile is the region below the tenth percentile, sometimes termed the first decile point. The tenth decile is the region above the 90th percentile or ninth decile point. The  $k$ th decile is the region between the  $10(k-1)$ th and  $10k$ th percentiles, or the  $(k-1)$ th and  $k$ th decile points.

The term "quartile" is often used also to represent a region. The lower quartile, the median, and the upper quartile are the three quartile points. The first quartile is the region below the lower quartile, the second quartile the region between the lower quartile and the median, the third quartile the region between the median and the upper quartile, and the fourth quartile the region above the upper quartile.

Centiles, deciles, and quartiles are *equal-frequency* score regions. Stanines and stens define equal standard score or normal-standard score regions, with unequal frequencies.

**Scaled scores.** A few intelligence and achievement tests and test batteries are designed to cover wide ranges of ability, e.g., grades 3-9 inclusive. More commonly, however, they are issued for successive levels, such as primary, elementary, advanced (junior high school), and in some cases secondary (senior high school) and adult. In achievement test batteries, additional subject areas are usually included at the higher levels; and at the primary level, picture tests may replace tests which at other levels require reading. The successive levels usually have similar materials differing in average difficulty, but with the harder items at one level overlapping in difficulty the easier items at the next higher level.

With wide-range tests and tests issued at several levels, grade scores, age scores, modal-age grade scores, and grade, age, or modal-age grade percentile ranks or standardized scores may represent quite unequal units at different levels. Scaled score systems are designed to have units which are equal in some sense, at least on the average, from level to level throughout the range. They are based on assumptions about the shape of the underlying ability distribution within a grade or age group, and are derived by considering both the score distributions at successive grades or ages and the

mean or median gains from grade to grade or age to age. None of these methods are wholly satisfactory.

Further problems arise when attempts are made to scale tests of different abilities in comparable units, since the relations between mean gains and within-grade or within-age variability are quite different for different functions. Thus, mean annual gain in reading is a much smaller fraction of within-group standard deviation than is mean annual gain in arithmetic; or, stated in terms of growth units, variability in reading is much greater than in arithmetic.

**Equating.** Before the scores on two tests, or even the scores on two forms of the same test, can be compared, the relations between their score scales must be established. The preferred experimental design is to give both tests to the same group of examinees: to half the group in the A-B order and to the other half in the B-A order.

The simplest method of establishing comparable scores is termed line-of-relation equating. Scores on test A and test B are considered comparable if they correspond to the same standard score. This method is satisfactory only if the item-difficulty distributions are of the same shape and are equally variable, which is seldom the case. The preferred method is termed equipercentile equating. Scores are considered comparable if they correspond to the same percentile. Selected percentiles are computed for each distribution, such as percentiles 1, 2, 3, 5, 10, 15, 20, 30, 40, 50, 60, 70, 80, 85, 90, 95, 97, 98, and 99. A two-way chart is prepared, with the scores on one form as abscissas and the scores on the other form as ordinates and the scores on the selected percentiles a point is plotted representing the scores on the two tests corresponding to this percentile, and a smooth curve is drawn as nearly as possible through all these points. If, but only if, this curve turns out to be a straight line, will line-of-relation equating have been satisfactory.

If the two distributions are first smoothed, the equipercentile points are more likely to lie on a smooth curve, and the accuracy of the equating is improved.

This method of equating is satisfactory if the two tests are equally consistent. If they are not, the scores on each test should all be multiplied by the consistency coefficient of that test, and the resulting "estimated true scores" should be equated.

When new forms of a test are issued annually, a full norms study is usually conducted only once every five or ten years, and the norms for successive forms are prepared by equating them to the "anchor form" used in the last norms study.

**Standards.** In a few cases, standards of test performance can be established without reference



to the performances of members of defined groups. Thus, in typing 120 words per minute with not more than one error per 100 words is a fairly high standard.

"Quality scales," in areas such as handwriting and English composition, are sets of specimens at equal intervals of excellence. A standard of handwriting legibility can be set by measuring the speed with which good readers read the various specimens. The standard would be the poorest specimen which is not read significantly slower than the best specimen. Units above this standard would then represent mainly increases in beauty; units below the standard, decreases in legibility. In English composition, the poorest specimen written in substantially correct grammar could be identified by a consensus of English teachers. Then units above the standard would represent mainly improvements in style; units below the standard, decreases in grammatical correctness.

Research is in progress to determine standards for multiple-choice tests of subject-matter achievement. When the items of such a test are arranged in order of actual difficulty, experienced teachers expert in the subject might be able to agree on the item which barely passing students should get right half the time. Given this item and the item-analysis data, the passing score for the test can be determined fairly readily.

**Expectancy tables.** When the regression of a test or battery on a criterion has been determined from a representative sample of the same population, norms for the population can be expressed in terms of expected criterion scores. The predictor scores are usually expressed in fairly broad units. Then, if the criterion is dichotomous, the expectancy table gives for each score level the probability that a person at that score level will be in the upper criterion group. If the criterion is continuous, the criterion scores are grouped into quartiles, deciles, stanines, stens, or grade levels; and for each predictor score level the table gives probabilities for the several criterion levels. Thus if the criterion is a grade in a course, the expectancy table will show for each predictor score level the probabilities that the grade will be A, B, C, D, or F.

EDWARD E. CURETON

[Directly related are the entries INTELLIGENCE AND INTELLIGENCE TESTING; MATHEMATICS; PERSONALITY MEASUREMENT; SCALING. Other relevant material may be found in FACTOR ANALYSIS; MULTIVARIATE ANALYSIS, articles on CORRELATION; NON-PARAMETRIC STATISTICS; STATISTICS, DESCRIPTIVE; and in the biographies of BINET; PEARSON; SPEARMAN; THORNDIKE; THURSTONE.]

## BIBLIOGRAPHY

*The standard work in psychometrics is Guilford 1936. It includes psychophysics and scaling, as well as the topics covered in this article. A somewhat more elementary treatment is given in Ghiselli 1964, and a somewhat more advanced treatment in Gulliksen 1950. Another general work is Stevens 1951, which treats with some care the foundations of measurement, a topic not covered by Guilford, Ghiselli, or Gulliksen.*

- BROGDEN, HUBERT E. 1949 A New Coefficient: Application to Biserial Correlation and to Estimation of Selective Efficiency. *Psychometrika* 14:169-182.
- CRONBACH, LEE J. 1951 Coefficient Alpha and the Internal Structure of Tests. *Psychometrika* 16:297-334.
- CRONBACH, LEE J.; and MEEHL, P. E. (1955) 1956 Construct Validity in Psychological Tests. Pages 174-204 in Herbert Feigl and Michael Scriven (editors), *The Foundations of Science and the Concepts of Psychology and Psychoanalysis*. Minneapolis: Univ. of Minnesota Press. → First published in Volume 52 of the *Psychological Bulletin*.
- FAN, CHUNG-TER 1952 *Item Analysis Table*. Princeton, N.J.: Educational Testing Service.
- FINDLEY, WARREN G. 1956 A Rationale for Evaluation of Item Discrimination Statistics. *Educational and Psychological Measurement* 16:175-180.
- FLANAGAN, JOHN C. 1939 General Considerations in the Selection of Test Items and a Short Method of Estimating the Product-moment Coefficient From the Data at the Tails of the Distribution. *Journal of Educational Psychology* 30:674-680.
- GHISELLI, EDWIN E. 1964 *Theory of Psychological Measurement*. New York: McGraw-Hill.
- GUILFORD, JOY P. (1936) 1954 *Psychometric Methods*. 2d ed. New York: McGraw-Hill.
- GULLIKSEN, HAROLD 1950 *Theory of Mental Tests*. New York: Wiley.
- HORST, PAUL 1936 Item Selection by Means of a Maximizing Function. *Psychometrika* 1:229-244.
- KUDER, G. F.; and RICHARDSON, M. W. 1937 The Theory of the Estimation of Test Reliability. *Psychometrika* 2:151-160.
- LORD, FREDERIC M. 1964 An Empirical Comparison of the Validity of Certain Formula-scores. *Journal of Educational Measurement* 1:29-30.
- LYERLY, SAMUEL B. 1951 A Note on Correcting for Chance Success in Objective Tests. *Psychometrika* 16:21-30.
- STEVENS, S. S. (1951) 1958 Mathematics, Measurement, and Psychophysics. Pages 1-49 in S. S. Stevens (editor), *Handbook of Experimental Psychology*. New York: Wiley. → See especially pages 13-15 and 21-30.
- Taxonomy of Educational Objectives*. Edited by Benjamin S. Bloom. 2 vols. 1956-1964 New York: McKay. → Handbook 1: *The Cognitive Domain*, by B. S. Bloom and D. R. Kratwohl, 1956. Handbook 2: *The Affective Domain*, by D. R. Kratwohl, B. S. Bloom, and B. B. Masia, 1964.
- THORNDIKE, EDWARD L. et al. 1926 *The Measurement of Intelligence*. New York: Columbia Univ., Teachers College.
- THORNDIKE, ROBERT L. 1951 Reliability. Pages 560-620 in E. F. Lindquist (editor), *Educational Measurement*. Washington: American Council on Education.
- WESMAN, ALEXANDER G. 1949 Effect of Speed on Item-Test Correlation Coefficients. *Educational and Psychological Measurement* 9:51-57.

## PSYCHONEUROSIS

See NEUROSIS.

## PSYCHOPATHIC PERSONALITY

The term *psychopathic personality* was for many years the officially approved designation of the American Psychiatric Association for a great many types of maladjustment in people who were otherwise regarded as free from psychosis, psychoneurosis, and mental deficiency. A large number of heterogeneous disorders, many of which have little in common, were included in this category. In 1952 the association officially replaced this term with the general category *personality disorders*. This category also includes many deviations, maladjustments, and disorders that vary widely in type and in degree. In the present scheme of nomenclature the subclassification *sociopathic personality: antisocial reaction* includes the type of patient well known to psychiatrists and for many decades referred to by the informal term *psychopath*.

**Definition.** The *Diagnostic and Statistical Manual, Mental Disorders* of the American Psychiatric Association defines antisocial reaction as follows:

This term refers to chronically antisocial individuals who are always in trouble, profiting neither from experience nor punishment, and maintaining no real loyalties to any person, group, or code. They are frequently callous and hedonistic, showing marked emotional immaturity, with lack of sense of responsibility, lack of judgment, and an ability to rationalize their behavior so that it appears warranted, reasonable, and justified.

The term includes cases previously classified as "constitutional psychopathic state" and "psychopathic personality." As defined here the term is more limited, as well as more specific in its application. (1952, p. 38)

Patients classified in this group are distinguished from other sociopaths, a category that includes those suffering from sexual deviation, alcoholism, and drug addiction; this group is also distinct from those exhibiting *dyssocial reaction*, which the *Manual* defines as follows:

This term applies to individuals who manifest disregard for the usual social codes, and often come in conflict with them, as the result of having lived all their lives in an abnormal moral environment. They may be capable of strong loyalties. These individuals typically do not show significant personality deviations other than those implied by adherence to the values or code of their own predatory, criminal or other social group. The term includes such diagnoses as "pseudo-social personality" and "psychopathic personality with asocial and amoral trends." (1952, p. 38)

It should be noted that the patients defined as showing dyssocial reaction are nonetheless capable of loyalties and seem in their rebellion against society to have some standards of their own, even though these may be immoral and condemned by law. This classification apparently covers the ordinary purposive criminal who works consistently toward antisocial but comprehensible goals.

This article will concern itself primarily with patients showing antisocial reaction, to whom we shall refer by the informal but now classic term *psychopath*. This term is familiar to all psychiatrists, and it clearly designates and distinguishes a well-known clinical entity.

Few modern definitions have surpassed J. C. Prichard's 1835 description of certain key aspects of the psychopathic personality:

... the moral and active principles of the mind are strongly perverted or depraved; the power of self-government is lost or greatly impaired and the individual is found to be incapable not of talking or reasoning upon any subject proposed to him, but of conducting himself with decency and propriety in the business of life. (Quoted by Cleckley 1959, p. 569)

Prichard is generally credited with formulating the concept *moral insanity*, which is still used in some parts of the world to designate the abnormality shown by psychopaths. In a recent study McCord and McCord give another definition that is helpful:

The psychopath is an asocial, aggressive, highly impulsive person who feels little or no guilt and is unable to form lasting bonds of affection with other human beings. (1956, p. 2)

## Characteristics of the psychopath

The psychopath does not give the impression, even on careful examination, of being mentally ill, handicapped, or emotionally disturbed. Nor does he, typically, show any attitude or outlook that would indicate he lacked conscience or had chosen rebellious or antisocial aims. His reasoning is excellent. What he tells of his allegiances, aims, and understandings indicates that he is normal, reliable, and utterly sincere. Despite this, his past conduct and what will emerge in the future are very likely to bear out the truth of Lindner's statement that here we encounter "the most expensive and most destructive of all known forms of aberrant behavior."

The typical psychopath is a person who appears to have at least average, and often unusual, ability and who seems to be clearly aware of the amenities and to affirm the moral code. Frequently he demonstrates superior intelligence and other assets and is likely to succeed brilliantly for a time in work,



in studies, and in all his human relations. But inevitably, and repeatedly, he fails, losing his job, alienating his friends, perhaps losing his wife and children. It is difficult to account for these failures. Seldom can one find adequate motivation to explain why a person has, in the midst of success, grossly shirked his immediate responsibilities, and perhaps abandoned his work, at the behest of impulses that seem to the observer no more compelling than a trivial whim. However effective he may show himself to be over a limited period, when given sufficient time, he proves himself inadequate. His failures deprive him of what he tells us are his chief objectives and also bring hardship, shame, and disaster to his wife, children, parents, and all those closely connected with him (Cleckley 1941).

In addition to such relatively passive types of failure most fully developed psychopaths also commit aggressive antisocial acts. They forge checks, swindle, steal repeatedly, lightly indulge in bigamy, and show little or no compunction about their sexual behavior, regardless of the consequences. Some psychopaths who have attracted wide public attention committed murder or other shocking felonies, usually with little or no provocation, often without comprehensible motivation (Cleckley 1941). The majority, despite many conflicts with the law, appear to avoid crimes sufficiently grave to result in their removal from society for long prison terms. The psychopath may repeatedly receive punishment that would be likely to cause an ordinary person to mend his ways. But he appears to learn nothing important from experience. He is quite familiar with the correct ethical criteria, claims allegiance to such criteria, and can formulate in words excellent rules and plans for himself to follow. He does not seem to be lying as simply as the ordinary liar, whose motives are usually comprehensible. Sometimes the psychopath does not seem to be aware that he is lying or even to grasp emotionally the essence of falsehood and how falsehood differs from truth. Sometimes such people seem to mean for the moment to do what they promise so convincingly, but the resolution passes almost as the words are spoken.

The psychopath expresses normal reactions (love, loyalty, gratitude, etc.) with a most impressive appearance of sincerity and depth, but the emotional ties and the attitudes he professes fail to deter him from deeds that continually contradict his verbal claims. There appears to be in him a strange lack of insight or, perhaps more accurately, a total lack of one of the dimensions that constitute insight. After innumerable lies that he knows have been detected, he still speaks confidently of giving

his word of honor, apparently assured that this will settle the issue immediately and absolutely. Although he may demonstrate, over considerable periods, adequate general abilities, or even extraordinary talent, he always throws away what he has gained, what he insists are his chief objectives. No adequate motivation can be found or even imagined to account for his conduct. Conceivable temptations are often extremely trivial, but they inevitably evoke actions that lead to the loss of fortune and the respect of friends, the destruction of marriage, and imprisonment or confinement in a psychiatric institution. The psychopath seems to be almost totally immune to real remorse or deep feelings of guilt or shame.

**An example.** At this point it is appropriate to consider the case of a young woman in her early twenties whose appearance of high intelligence is confirmed by psychometric examination and by excellent achievement in her studies and at work during the limited, sporadic periods during which she applied herself. Despite this intelligence she has been a serious problem to her parents since early childhood. She has repeatedly stolen from her family, from strangers, and from stores. Often she has bought unneeded clothes and other articles and has habitually lied with equanimity and effectiveness. At times she has run up such large, unauthorized bills that her father found himself in serious financial difficulties. No punishment or deprivation has influenced her in the slightest degree. As a girl she had been frequently truant from school and violated every agreement made about suitable hours to return from dances or dates. Yet she has almost never been openly or directly defiant and has usually admitted herself in the wrong when her cool and ingenious falsehoods were exposed. She regularly promises, in a convincing and apparently sincere manner, to reform completely but after varying periods of good behavior always returns to her old patterns of irresponsibility and folly.

She has had a number of boy friends but apparently feels less than the average sexual interest. Often she has seemed to be mildly bored but seldom distinctly unhappy. During the year before her admission to the hospital she spent hours each week writing sentimental and rather perfunctory letters, chiefly to boys in military service or at college. Some of these she knew only slightly. When one of the boys, whom she knew much better and actually considered her sweetheart, was killed in an automobile accident, she manifested little sign of grief or concern. She remained a virgin until she was 18 years old. Then one day, without apparent

provocation or strong incentive, she left home secretly and was not heard from until two weeks later when police located her in a city approximately a hundred miles away. She showed no remorse about the anxiety her parents had suffered and neither disillusionment nor resentment about the fruitlessness of her venture or the rather brutal treatment she had occasionally experienced.

In explanation of her conduct she said that she had left home with the intention of visiting a boy friend stationed at an army camp in another state. She admitted that she had in mind the possibility of marrying this man, but that no final decision had been made by her, much less by him. She admitted that she had given the matter little thought, and from her attitude one would judge she was moved by little more than a casual whim.

She had left with only a few dollars in her purse. Getting off the bus in a town fifty miles from home, she tried to reach the boy by telephone and ask him to telegraph funds to her. She could not reach him and decided to go to the home of another boy, then overseas, whose family lived in the town where she found herself. She asked these people to let her spend the night. They welcomed her. After being there an hour or two, she became suspicious that they might have notified her parents. Without a word to them she slipped out of the house, took a bus in another direction and, late at night, got off in a small, strange town.

For a little while she sat in a hotel lobby, apparently calm and unworried. Soon a man approached her and drew her into conversation. On hearing her story, he offered to pay for a room so she could rest. She realized he meant to spend the night with her, but she made no objection. As far as one could tell by discussing the experience with her, she was neither alarmed, excited, repulsed, nor attracted by a prospect that most carefully raised virgins could not have regarded with indifference. The man apparently handled her in a crude, inconsiderate fashion and showed plain contempt for her and her role, addressing her in unnecessarily derogatory terms. During sexual relations with him she experienced moderate pain and no sexual excitement or response. He gave her a few dollars, which she was pleased to accept.

The next day she telephoned her soldier friend at his camp. She had not discarded her idea of perhaps marrying him, nor had she progressed any further toward a firm decision to do so. He discouraged her strongly against coming, refused to send her money, and emphatically urged her to return home. Apparently she was not upset by this turn of events and, with little consideration of the

matter, decided to go to the city where she was eventually discovered. She admits that she had no definite purpose in going there, was prompted by no overmastering thrill of adventure, and was not aware of any strong reason for not returning home or for having left in the first place. In the strange city she got small jobs for a day or two, but these did not pay enough for room and board. She therefore began to spend the nights with various men in search of light sexual adventure. She experienced no real erotic enjoyment from these experiences or any serious qualms of conscience. When detected by police and, later, on meeting her friends at home, she never showed embarrassment or any sign of remorse about what had occurred.

Somewhat similar episodes recurred time after time. On various occasions she expressed the wish to go into training as a student nurse, to do secretarial work, or to get a job in a government agency. At considerable expense and trouble her family enabled her to try all these plans and many others. Each time, she expressed the most confident assurance that she would have no further trouble. In each new venture she made a favorable impression at first and showed herself capable of succeeding and often of excelling. Soon, however, she abandoned each position, sometimes running away to other towns and living more or less as a vagabond. Once she absconded with money obtained from a lending agency operated by the corporation she worked for. She often stole things for which she had no use. Sometimes without any particular reason she simply quit going to the office and in this way lost good positions in which she had expressed great interest and at which she was succeeding with minimal effort.

After dozens of such failures she continued to express the most complete optimism, and her attitude always implied buoyant self-satisfaction.

After coming to the hospital a number of times for psychiatric observation and treatment, she wrote to her physician:

You have given me a new outlook and a new life; shown me things in a different light—and I am happy—and have every hope of making those who, you could even say, have *saved* me, proud of me. Whether I turn out finally to be a nurse, a business girl or just a wife, I have a new hope for a happy and useful life which I thought would never be mine. If in this new life I ever become discouraged and lose hope I will look at you and have courage and hope again. I now understand myself and my problems, thanks to you. I will go on and succeed and be happy.

After several prompt repetitions of previous failure she continued to give her psychiatrist's name



as a recommendation for positions on the basis of her "good character," "reliability," "high moral standards," etc. She seemed perfectly assured that such recommendations would be in the highest terms and without qualification.

It is not easy to fathom the motivation of such a person. If in hallucinations she had heard God's voice urging her to leave home or had believed she was being invited to spend the night by a famous motion picture star or a princely lover, her conduct would be easier to understand; it would, in an important sense, be less unreasonable. It would also be easier to account for if she had been tempted by strong and persistent sexual desire or some other major incentive to sacrifice her good name and perhaps her chances of future happiness for a foolish but fascinating or wildly exciting adventure.

**Distinguishing characteristics.** The patient whose history has just been briefly summarized shows many of the characteristics that seem to distinguish the real psychopath from other people, including other psychiatric patients. The following is a list of some of the characteristics that have emerged in the study of many patients of this type: (1) Easy charm and good technical intelligence. (2) A confident and forthright manner that inspires trust. (3) An apparent dignity and seriousness that give the impression of profound sincerity. (4) The absence of delusions, hallucinations, and all other signs of psychotic or irrational thinking. (5) Freedom from abnormal anxiety and other manifestations of psychoneurosis. (6) Ability to formulate excellent life plans that seem to indicate sound judgment and steadfast resolution. (7) Allegiance to moral codes and conventions of society, expressed cogently and with the appearance of deep conviction.

These points are likely to be impressive during psychiatric examination or in any other encounter confined to the patient's verbal performance as contrasted with his behavior. When we study his behavior over a considerable period of time, other points emerge that are strikingly incompatible with those listed above: (1) Unreliability and irresponsibility in behavior even when very important issues are at stake and personal welfare is involved. (2) Lack of remorse or shame despite eloquent protestations of such feelings. (3) Inadequately motivated antisocial behavior repeated again and again despite interludes of good adjustment. (4) Conduct that demonstrates extremely poor judgment and failure to learn by experience. (5) Inability to love another, and pathologic egocentricity. (6) Shallowness and poverty in the major human emotions. (7) A specific loss of insight despite

ability to use words that convincingly give the opposite impression. (8) Unresponsiveness to the important issues that arise in interpersonal relations. (9) Uninviting and outlandish behavior, often with little discernible provocation. (This is particularly prone to emerge when drinking.) (10) Tendency to threaten suicide frequently but rarely to attempt it, despite circumstances that would make strong impulses toward suicide readily understandable. (11) Shallow, impersonal, trivial, and poorly integrated sex life. (12) Failure to follow consistently any life plan. (Although brilliant successes occur, the results of these are often dissipated or destroyed on impulses that appear to be little more than caprice.)

### Nature of the disorder—etiology

The disorder of the psychopath varies considerably in degree, and some of these patients, despite their technical sanity, prove themselves much less competent, less able to live their lives effectively or without serious damage to themselves and to others than people who show plain manifestations of psychosis. The psychopath, although entirely free of irrationality and often brilliant in all verbal tests, carries on a career that is difficult to account for as a product of sanity.

What is the nature of the disorder or defect in these people? Since the mid-nineteenth century various answers have been offered to this question. It has been said that they suffer from pathologic changes confined to a hypothetical "moral faculty." It has also been said that they are affected by general constitutional inferiority, subtle hereditary taint, the deprivations of slum life, the influence of evil companions, and parental neglect. Some of these explanations have probably emerged from the observation of patients now distinguished as showing dyssocial reaction and of other types of disorder differing fundamentally from that of the true psychopath but arbitrarily placed in the same category. During recent decades increasing attention has been paid to the influence of environmental or psychogenic factors rather than possible inborn or organic defects. Some observers have expressed the conviction that the psychopath is a product of extreme parental spoiling. Others have seemed as confident that early and subtle rejection by the parents is a major factor in his development (Cleckley 1941).

**Franz Alexander's formulation.** Since the 1930s the concepts of Franz Alexander have attracted much attention and have, indeed, been accepted by many psychiatrists as furnishing a scientific explanation of the problem. According to Alexander's

formulation, the psychopath's maladjustment and antisocial behavior arise from inner conflict similar to that widely believed to underlie psychoneurotic symptoms. Despite the striking lack of ordinary anxiety or self-condemnation in psychopaths, even under circumstances in which these reactions would be normal, Alexander maintains that there is an inner unconscious conflict and sense of guilt (1930).

In reaction to such a conflict the patient with psychoneurosis is thought to develop subjective symptoms and to complain of anxiety, headache, weakness, and other physical disturbances or perhaps to become disturbed by compulsive rituals, obsessive thinking, hysterical paralysis, and temporary blindness. Alexander maintains that the psychopath, instead of passively developing such unpleasant subjective symptoms, reacts to the postulated unconscious conflict by indulging in destructive, self-damaging, antisocial behavior. This process is thought of by Alexander as an "acting out" of the inner problems in relation to society and to the environment (1930). Many of those who accept this hypothesis believe that the psychopath deliberately but unconsciously seeks failure and persistently commits crimes and other acts of aggression to obtain punishment, which he unwittingly desires as an expiation of intense but completely unconscious feelings of guilt and remorse. In this interpretation the psychopath can be thought of as having genuine and adequate reasons for the apparently foolish, antisocial, and uncalled-for acts he commits, damaging himself and others. It is assumed that he knows nothing of these reasons or of their effect upon him. This is in many respects an ingenious formulation. It is widely regarded as offering a "dynamic" explanation of the psychopath. It tends also to make him a dramatic figure and to arouse sympathy for his alleged blind struggle for redemption. [See ALEXANDER.]

**Adelaide Johnson's formulation.** Another and more recent explanation has been offered by Adelaide Johnson (1959). She also maintains that there are factors operating in the unconscious, not only of the patient but also of his parents. According to this formulation, the parents have criminal impulses of which they remain completely unaware and, without realizing it in the least, work deliberately and persistently to influence their son or daughter to carry out irresponsible, immoral, and destructive antisocial activity in order to obtain for themselves a vicarious satisfaction they do not dare seek more directly. According to this theory, the child, even after he has reached adult life, does not

become conscious of this alleged motivation or of the beguiling influence said by Johnson to be exercised unwittingly by the parents. Through this purposive training, the child is said to be inducted by the parent into a career of unrewarding delinquency and self-defeating antisocial behavior.

**Evaluation of these formulations.** Johnson's hypothesis, like Alexander's, seems to offer a "dynamic" explanation in the popular terms of unconscious motivation. Perhaps for this reason both have attracted much attention and have been widely accepted. They are regarded by many as scientific discoveries based on demonstrable evidence.

Let us not forget that both of these intriguing theories rest entirely on arbitrary assumptions about what is in a hypothetical unconscious. None of the patients I have studied have shown anything that could reasonably be regarded as evidence of the hidden sense of guilt and the desperate quest for redemption through punishment attributed to psychopaths by Alexander. Nor have I found anything to indicate that their parents wanted them to indulge in misconduct and, without realizing it, actively influenced them toward it. Such motives and feelings, or any others, may of course be assumed to be present not only in psychopaths but in anybody else if we also assume that they remain unconscious. Care, however, must be taken to distinguish between assumption and proof. Popular methods in psychiatry and psychology generally regarded as dynamic make it easy to ascribe with confidence virtually anything to the unconscious and to obtain pseudoevidence for whatever one might seek to establish, using analogy and the arbitrary interpretation of dreams and symbols. The psychiatric literature abounds in examples of these methods being so used to discover specious explanations of psychiatric disorder in events imagined by the psychiatrist, in accordance with his theory, to have occurred decades earlier, during the first few months of a patient's life; or, indeed, even during the patient's intra-uterine existence as an embryo (Cleckley 1957).

Perhaps some people carry out criminal acts repeatedly because of guilt they do not know they experience and do so in order to obtain punishment they do not know they seek. Perhaps there are law-abiding and respected parents who unconsciously want their children to rebel against society and indulge freely in forgery, burglary, bigamy, prostitution, and other forms of misconduct. However, no evidence has emerged in my own experience with psychopaths to indicate that such influences are likely to play a part in their disorder.



Until real evidence, as contrasted with mere assumption, can be offered to establish the concepts of Alexander and Johnson, let us regard them with skeptical interest.

For a long time it has been the custom to assume that in unconscious conflict can be found the explanation of psychiatric disorder, of delinquency, and, indeed, of human behavior in general. Jenkins (1960) suggests that it may be worthwhile for us to consider possible effects of the opposite, of a lack of conflict under circumstances that would normally cause anxiety and conflict to develop and that would actually make this imperative. He very pertinently points out that it became known early in the history of medicine that what a person ate could produce illness. Only much later did it become known that what he did not eat might cause a more subtle illness through avitaminosis.

**Masked personality disorder.** Is there some defect or disorder within the psychopath that causes him to lack the capacity to feel guilt? If so, this hypothetical deficiency seems also to interfere with his reacting to, and pursuing consistently, the normal goals of life. And he seems to lack the ability to participate adequately in the major emotional experiences of life.

The typical psychopath's excellent intellectual abilities and his freedom from the manifestations of ordinary psychiatric disorder make it difficult to believe that deep within him may be concealed a deficiency that leads not to conflict or unconscious guilt but, instead, makes him incapable of feeling normal remorse and of appreciating adequately the major emotional experiences of human life (Cleckley 1941).

The outer characteristics of the psychopath strongly indicate warmth of feeling, kindness, sincerity, pride, courage, a deep sense of honor, and genuine capacities for love and loyalty. Such an outer appearance could be the result of excellent peripheral function in the organism, which gives strong and convincing promise of robust health within and makes it difficult to suspect that there may be a central and very serious inner defect. The psychopath's conduct, however, is consistent with a serious defect in the very qualities for which his superficial aspect and verbal performance give such rich promise. The peripheral mechanisms, one might say, of his functional entity are undamaged and operate well. They demonstrate technical intelligence and convincingly mimic the expression of normal inner experience. But the implied inner experience, the glowingly promised emotional participation in life, is not there.

If we compare speech disorders with personality disorders, an analogy emerges that may be helpful

in conveying this concept. When the outer physiologic apparatus involved in the production of speech is damaged, the disability is overt, and its cause is usually easy to understand. When the tongue is mutilated or its motor nerve damaged, there is likely to be gross difficulty in enunciating words and perhaps even in moving the tongue itself. Efforts to speak may give rise only to inarticulate sounds that communicate nothing. The inner use of language, however, and its meaning to the person who has suffered the injury, is preserved intact. In contrast with dysarthria, in which the peripheral apparatus of speech is affected, let us consider the aphasias which are caused by lesions more centrally located in the brain. In these the outer mechanisms of speech are preserved.

Let us consider particularly semantic aphasia as described by Henry Head (1926). In this very deep-seated disorder of speech, words are clearly and accurately enunciated, and often complete and grammatical sentences are fluently spoken. These utterances, however, have little or no meaning. They are not related, within the person, to ideas or feelings that they seem to indicate and seem intended to convey. The words of this ostensible communication are, in a very important sense, not really words but only a mimicry of words, produced mechanically by the peripheral mechanisms of speech that have become isolated from the inner source that gives rise to thought, feeling, and intention. Despite this more or less reflex simulation of real speech, a deep loss has occurred that prevents the person from using language inwardly to think. [See LANGUAGE, article on SPEECH PATHOLOGY.]

If the psychopath has a profound and centrally located defect that prevents him from participating significantly in man's deepest fulfillments and joys, is it not possible that this inability to participate might contribute to restlessness and boredom? And might this not in turn prompt him to indulge in unprofitable or damaging indiscretions and destructive behavior that would not be particularly tempting to others who are devoting their attention to major goals and responding to major fulfillments? This hypothesis—of an extremely serious central pathology or a biological deficit concealed by misleading peripheral functions, by what one might call an impressive "mask of sanity"—cannot be established by objective evidence at present but is, in many important respects, consistent with the psychopath's behavior (Cleckley 1941).

It may be helpful to consider the case of patients who are affected by what has often been called "masked schizophrenia." Though these patients do not have delusions or hallucinations and are often

rational in their verbal expression of thought, the psychosis may be as genuine and its degree as great as in the hebephrenic patient, whose extremely disabling disorder is obvious. The patient with masked schizophrenia, unlike the psychopath, nearly always shows on examination a brittleness, some undefinable peculiarity of manner, an emotional coolness, or, perhaps, certain subtleties of posture, gesture, expression, or attitude that indicate he is by no means normal. It might be said that his psychosis is concealed but that it is concealed by the appearance of lesser abnormality, by the outer disguise of a very eccentric and peculiar person. In contrast to this, the mask of the psychopath is that of complete and robust health.

### Treatment

Over the years most observers have been impressed by the failure of psychopaths to respond to any treatment or be changed fundamentally by any other influence (Cleckley 1941). A few psychiatrists have expressed encouragement about psychoanalytic treatment of patients they regarded as psychopaths (Johnson 1959; Karpman 1955; Schmideberg 1949). Lindner enthusiastically reported success with the use of hypnoanalysis (1944). Others have occasionally reported improvement through the use of drugs, through the effects of simple counseling, through milieu therapy, through the methods of general semantics, and through various other types of psychotherapy (Lipton 1961; Lynn 1938; McCord & McCord 1956; Thompson 1953). Occasionally psychiatrists have been favorably impressed by reports of the use of electric shocks (Darling 1945) and lobotomy operations (Banay & Davidoff 1942). Perhaps some of these optimistic reports, which conflict so sharply with the experience of most observers, can be accounted for by the psychopath's tendency to use his excellent abilities temporarily—and perhaps one might also say, capriciously—to succeed for a while in whatever he undertakes. These spontaneous remissions, like those long familiar in multiple sclerosis, may lead the therapist to an erroneous belief that they are the result of his treatment. Most psychiatrists share the opinion that there is no effective treatment available today for genuine and typical psychopaths [see MENTAL DISORDERS, TREATMENT OF; see also Cleckley 1941; Greenacre 1947].

### Proposals for dealing with the problem

It would seem that important steps can be taken toward working out a better means of dealing with the many problems posed by the psychopath. First, a more general recognition is needed that his dis-

order is a genuine incompetency or psychiatric abnormality, a specific pattern of malfunctioning that, in severe cases, causes maximal disability and makes it impossible for him to maintain his role as a responsible member of the community. At present most psychopaths elude restriction. Neither penal institutions nor psychiatric hospitals possess the legal means of holding them long enough to afford the community protection or to give an adequate trial to any therapeutic measures that may be attempted.

Although the psychopath's record of extreme incompetency may facilitate his efforts to obtain exoneration or leniency in the courts when faced with imprisonment, it does not, as matters now stand, enable society to keep him under psychiatric supervision through commitment. Scores of times patients repeat apparently purposeless thefts, forgeries, embezzlements, bigamies, swindlings, distasteful or indecent acts in public, and yet, because the official medical category in which they are placed makes them sane and competent according to the books and tradition, they cannot be committed to psychiatric institutions for medical care or for the protection of themselves and others. If by chance they are committed, which is exceptional, their "sanity" is soon re-established by a staff of able psychiatrists at the institution to which they are sent, who correctly diagnose them as cases of psychopathic personality. However genuine their real disorder, this frees them (as legally competent persons) from control or supervision.

On the other hand, these patients frequently, and in most courts usually, are able to evade prison sentences for their antisocial acts. Their lawyers are able to point out the obviously incompetent features in their careers, and the jury, despite expert psychiatric testimony to the contrary, is often unwilling to punish persons whose conduct shows such plain evidence of mental abnormality. The Durham Rule—that the accused is absolved from criminal responsibility if his act was a product of mental disease—has apparently not resulted in any improvement in this confused situation [see PSYCHIATRY, article on FORENSIC PSYCHIATRY].

If a means could be made available of obtaining adequate control over psychopaths who plainly show themselves not fitted for unrestricted freedom in the social group, it would then be possible to set up facilities specifically designed to deal with their problems. Large state and federal psychiatric institutions in the United States, organized for the treatment of patients psychotic in the traditional sense, are not at present well adapted to handle the psychopath. Numerous private hospitals, whether primarily designed for the needs of the psycho-



neurotic patient or for the psychotic, lack the means of restraining the psychopath and are unable to deal adequately with the problems he creates. There would be no use in establishing institutions primarily for psychopaths unless legal means of controlling them were made available. The expense of building and operating specialized institutions of this sort would be great. There is good reason to believe, however, that even this would cost less than what the psychopath is costing the public today. And even if no effective treatment should ever be discovered, systems of parole and probation specifically set up for the psychopath and carefully designed in the light of his needs and his proclivities could give each patient as much freedom from restraint as he showed himself capable of utilizing safely. In a properly and persistently controlled situation it might be possible for many psychopaths to use their excellent abilities constructively and eventually achieve a much better adjustment than is possible for them today. If laymen were to become more acquainted with this enigmatic figure behind his misleading mask of sanity, and with the ever-accumulating sorrow, damage, despair, confusion, farce, and disaster that each psychopath leaves in his wake, organized effort might be mobilized to devise adequate medicolegal means of restraining him in his now virtually unhindered career of folly and purposeless destruction.

HERVEY M. CLECKLEY

[Directly related are the entries CHARACTER DISORDERS; CRIME; DELINQUENCY; DEVIANT BEHAVIOR. Other relevant material may be found in MENTAL DISORDERS; MORAL DEVELOPMENT.]

#### BIBLIOGRAPHY

- ALEXANDER, FRANZ 1930 The Neurotic Character. *International Journal of Psycho-analysis* 11:292-311.
- AMERICAN PSYCHIATRIC ASSOCIATION, COMMITTEE ON NOMENCLATURE AND STATISTICS (1952) 1963 *Diagnostic and Statistical Manual: Mental Disorders*. Washington: The Association.
- BANAY, RALPH S.; and DAVIDOFF, L. 1942 Apparent Recovery of a Sex Psychopath After Lobotomy. *Journal of Criminal Psychopathology* 4:59-66.
- CLECKLEY, HERVEY M. (1941) 1964 *The Mask of Sanity*. 4th ed. St. Louis, Mo.: Mosby.
- CLECKLEY, HERVEY M. 1957 *The Caricature of Love*. New York: Ronald.
- CLECKLEY, HERVEY M. 1959 Psychopathic States. Volume 1, pages 567-588 in *American Handbook of Psychiatry*. Edited by Silvano Arieti. New York: Basic Books.
- DARLING, HARRY F. 1945 Shock Treatment in Psychopathic Personality. *Journal of Nervous and Mental Disease* 101:247-250.
- GREENACRE, PHYLLIS 1947 Problems of Patient-Therapist Relationship in the Treatment of Psychopaths. Pages 378-383 in Robert Lindner and Robert V. Sel-

- ger (editors), *Handbook of Correctional Psychology*. New York: Philosophical Library.
- HALL, JEROME 1963 The M'Naghten Rule and Proposed Alternatives. *Journal of the American Bar Association* 59:960-964.
- HEAD, HENRY 1926 *Aphasia and Kindred Disorders of Speech*. 2 vols. Cambridge Univ. Press.
- JENKINS, RICHARD L. 1960 The Psychopathic or Antisocial Personality. *Journal of Nervous and Mental Disease* 131:318-334.
- JOHNSON, ADELAIDE M. 1959 Juvenile Delinquency. Volume 1, pages 840-856 in *American Handbook of Psychiatry*. Edited by Silvano Arieti. New York: Basic Books.
- KARPMAN, BENJAMIN 1955 Criminal Psychodynamics: A Platform. *Archives of Criminal Psychodynamics* 1:3-100.
- LINDNER, ROBERT 1944 *Rebel Without a Cause*. New York: Grune.
- LIPTON, HARRY R. 1961 The Psychopath. *Archives of Criminal Psychodynamics* 4:542-549.
- LYNN, JOHN G. 1938 Preliminary Report of Two Cases of Psychopathic Personality With Chronic Alcoholism Treated by the Korzybskian Methods. Pages 90-100 in American Congress of General Semantics, First, Ellensburg, Washington, 1935, *General Semantics: Papers*. New York: Arrow.
- MCCORD, WILLIAM; and MCCORD, JOAN 1956 *Psychopathy and Delinquency*. New York: Grune.
- PRICHARD, J. C. 1835 *Treatise on Insanity and Other Disorders Affecting the Mind*. London: Gilbert & Piper.
- SCHMIDBERG, MELITTA 1949 Psychology and Treatment of Criminal Psychopaths. Pages 174-189 in K. R. Eissler (editor), *Searchlights on Delinquency*. New York: International Universities Press.
- THOMPSON, GEORGE N. 1953 *The Psychopathic Delinquent and Criminal*. Springfield, Ill.: Thomas.

## PSYCHOPATHOLOGY

See MENTAL DISORDERS.

## PSYCHOPHARMACOLOGY

See under DRUGS.

## PSYCHOPHYSICS

Psychophysics is the study of physical stimuli and their relation to sensory reactions. Of the energy that strikes the sensory surfaces of man and the animals, only a restricted fraction is capable of eliciting a reaction. Thus, visual responses in man are triggered by a narrow band of the vast electromagnetic spectrum (wavelengths between about 400 and 750 millimicrons); and auditory responses result from periodic displacements of the eardrum in the frequency range from about 20 to 20,000 cycles per second (cps). Over these stimulus ranges, neither the eye nor the ear is uniformly responsive: to produce a sensation may require thousands of times more energy at one wavelength or frequency than at another. It is the goal of psychophysics to map out the relations between the physical events and the psychological re-

sponses of organisms, and thus to provide a basic, over-all description of the function of the senses.

Major problems. The traditional questions posed by psychophysics fall into four groups. For a given sense modality we may ask about (1) the smallest detectable energy (the measurement of *sensitivity*); (2) the smallest detectable change in energy (the measurement of *resolving power*); (3) the configurations of energy that produce an invariant sensory effect, such as a constant loudness or color (the measurement of *static invariances*); (4) the way in which the magnitude of a sensory effect depends functionally on the stimulus (the measurement of *dynamic properties*).

### Sensitivity

The problem of sensitivity involves the determination of the smallest detectable intensity of a stimulus (called the *absolute threshold*), often as a function of another stimulus dimension, such as wavelength, frequency, duration, or areal extent. The threshold of audibility, for example, depends on the frequency of the tone; sensitivity is greatest to frequencies between about 2,000 and 3,000 cps. In the vicinity of 20 and 20,000 cps (which are conveniently but arbitrarily called the "limits" of hearing), the threshold energy may rise to roughly  $10^6$  times the minimal value. A similar relation exists between the threshold of visibility and the wavelength of the stimulating light, except that the shape of the visibility curve depends on what part of the sensory surface is stimulated; foveal (cone) vision has maximum sensitivity at about 555 millimicrons; peripheral (rod) vision, at about 505 millimicrons; and the "limits" of visibility are also different for the two populations of receptors.

Variation in sensitivity. The absolute threshold is not a rigidly fixed value. Sensitivity fluctuates irregularly, so that a given stimulus level may trigger a response at one time but not at another. The threshold is usually defined statistically, e.g., as the energy level that is detected as often as not over a series of presentations.

Sensitivity is also subject to systematic variation, either of the permanent kind encountered in aging or in pathology of the sensory tissues, or of a temporary kind observed, for example, in the relatively rapid decline of visual sensitivity under exposure to light (light adaptation) and the subsequent gradual recovery of sensitivity in the dark (dark adaptation). These and many other systematic changes in sensitivity are frequently expressed as alterations of the absolute threshold.

The study of absolute thresholds reveals the exquisite sensitivity of the sense organs under optimal conditions. A periodic displacement of the ear-

drum through a distance equal to the diameter of a hydrogen molecule may suffice to produce an audible sound, and a couple of quanta of light absorbed at the retina may suffice to arouse a faint visual sensation.

Methods of measurement. Because it fluctuates, the threshold is difficult to measure, and the various methods that have been tried do not always yield the same value. The method of *adjustment* provides a rapid approximation; the observer is required to set the level of the stimulus so that it is just perceptible. The threshold may be defined as the average of several settings. In the method of limits, either the stimuli are presented in order of increasing magnitude until the observer reverses his response from "imperceptible" to "perceptible," or they are presented in order of decreasing magnitude until the observer reverses his response from "perceptible" to "imperceptible." The threshold may be defined as the average value that marks the reversal in response over several ascending and descending series. In the method of *constant stimuli*, fixed stimulus levels are presented several times, each in irregular order. The threshold may be defined as the stimulus value that is perceived on half the presentations. This value is interpolated from a plot relating the percentage of positive responses to the stimulus magnitude.

The methods of adjustment, limits, and constant stimuli are known as the *classical psychophysical methods* because they have continued in widespread use ever since G. T. Fechner described them in his *Elemente der Psychophysik* (1860), the monumental work that marks the establishment of psychophysics. (Reviews of the classical methods are given by Urban 1908; Titchener 1905; and Boring 1942.)

One of the difficulties inherent in these methods is the observer's awareness that a stimulus event actually takes place on each trial. When a "catch trial" (a feigned presentation of a stimulus) is given, observers will occasionally give an affirmative response (a "false alarm"). The knowledge that the observer's expectations and motivations come into play has stimulated the invention of new methods that offer the hope of better understanding and controlling the observer's response biases. An example is the *forced choice* method, in which at regular intervals a stimulus is presented or withheld and the observer must decide each time whether or not he detected it. Results obtained under this procedure reveal that detection may depend not only on the magnitude of the stimulus but also on the prearranged probability of a stimulus event. A high proportion of "no-stimulus" trials causes a relatively high incidence of "false alarms";



a low proportion of "no-stimulus" trials, on the other hand, causes a lower incidence of correct detections of actual stimulus events. The probability of a "Yes" or "No" response can also be systematically influenced by rewarding correct detections and punishing the false alarms.

The forced-choice experiments have done much to underscore and clarify the role of response variables in the measurement of thresholds. It is sometimes suggested that the "detection" model may actually do away with the conception of the threshold as a simple, determinable value marking the critical terminus of sensory experience. According to this view, the detection of a stimulus (the "signal") has much in common with the mathematical process of statistical decision. The observer is confronted with two distributions: that of the persistent background noise and that of the signal added to the noise. He decides from which of the two distributions a sample is taken in much the way that a statistician tests a statistical hypothesis. The decision will depend on the overlap of the distributions and also on the "pay-off matrix"—the consequences of false detections and failures of detection (see Swets 1964; Luce et al. 1963).

Interesting technological advances have recently been made in the field of threshold measurement. The Békésy audiometer, for example, uses the method of *tracking* for the efficient measurement of the just-audible intensity as a function of tonal frequency (Von Békésy 1928–1958). The observer "tracks" his threshold by pressing a key whenever the tone is audible and releasing it whenever the tone becomes inaudible. While the key is pressed, the level of the tone steadily decreases; while the key is not pressed, the level steadily increases. The observer may continue to track the threshold while the tonal frequency changes from one end of the audible spectrum to the other. On a moving paper chart, the stimulus level, which weaves back and forth across the threshold, is recorded continuously as a function of the frequency.

The tracking method has also been used to determine the visual thresholds of human observers and has been adapted to mapping the sensitivity functions of animals.

### Resolving power

The second major concern of psychophysics is to measure the smallest detectable change in a stimulus (the so-called *difference threshold*). The problem may be to measure the just-noticeable differences in *intensity*, e.g., in the brightness of a light or in the concentration of a sweet solution, or in *quality*, e.g., in the hue of a colored light.

The capacity for resolving stimulus differences is expressed in terms of the Weber fraction,  $\Delta I/I$ , where  $\Delta I$  stands for the increment that produces a just-noticeable change when added to the stimulus level  $I$ . The smaller the value of  $\Delta I$ , the keener is the ability to discriminate.

**Methods of measurement.** Like the absolute threshold, the difference threshold is a fluctuating quantity, so that  $\Delta I$  must be assessed by a statistical treatment of a series of measurements. Most of the methods used are versions of those used to measure absolute thresholds. In the method of *adjustment*, for example, the observer sets a comparison stimulus to match a standard fixed stimulus. The threshold,  $\Delta I$ , may be defined as the average error or the standard deviation of several settings. The greater the variability of the settings, the grosser the discrimination and the larger the Weber fraction. A difference threshold may be regarded either as a measure of the precision or as a measure of the variability or "noisiness" of the sensory process.

In the method of *constant stimuli*, the observer judges whether each of a set of fixed discrete stimulus levels appears greater or smaller than a standard stimulus (a judgment of "equal" is also permitted by some experimenters). The difference threshold may be defined as the difference between a standard stimulus and a comparison stimulus that is perceived as being greater (or smaller) than the standard stimulus on a certain percentage of the trials. This value can be interpolated from a *poikilitic* (scatter) *function*.

In Figure 1 the ordinate represents the relative frequency with which the comparison stimulus is judged greater than the standard. The threshold,  $\Delta I$ , is the difference between the stimulus magnitude that is perceived as being greater than the standard on 75 per cent of the trials ( $L$ ) and the stimulus magnitude that is so perceived on 50 per cent of the trials ( $E$ ), i.e., the stimulus that appears to match the standard stimulus. Often there is a small difference, called the time error, between the standard stimulus ( $S$ ) and the stimulus value associated with the 50 per cent point. Urban (1908) provides a detailed discussion of poikilitic functions.

The nature of the difference threshold has often been studied by the method of *quantal increments*. From time to time, a small increment,  $\Delta I$ , is added briefly to a steady stimulus,  $I$ . The task is to indicate whether the increment was detected. Of theoretical concern is the mathematical form of the poikilitic function that relates the proportion of detections to the size of the increment. If the precision were ultimately limited by nothing but the

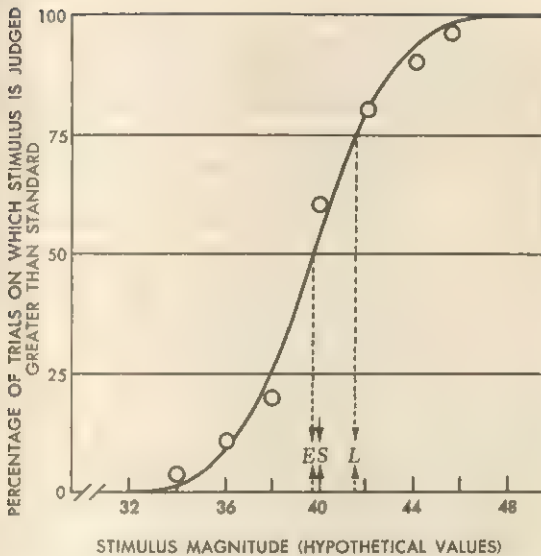


Figure 1 — A poikilitic (scatter) function determined by the method of constant stimuli

random "noisiness" of the sensory process, then the function would be expected to have the sigmoid shape (the integral of a bell-shaped distribution) that is predicted by the theory of random error. The usual result approximates this form. When pains are taken to aid the attention and to eliminate the extraneous sources of error, the obtained function may assume a linear rather than a sigmoid form and conform to a predictable slope. According to the *neural-quantum hypothesis*, a linear function of the appropriate slope demonstrates that discrimination is basically all-or-none and that sensation grows by the addition of minute but finite (or quantal) steps (see Von Békésy 1928–1958; S. S. Stevens 1961; Luce et al. 1963; Swets 1964).

**Weber's law.** A major aim in threshold measurement has been to test the famous generalization, credited to Ernst H. Weber but formalized and promoted by Fechner, that the Weber fraction is constant along a given sensory continuum, or that  $\Delta I/I = k$ . In other words, a just-noticeable change should occur when a constant fractional increment is added to a stimulus of any magnitude. Under good conditions the increment is about 1 per cent for brightness, 2 per cent for loudness, and 20 per cent for saltiness. The sensory systems differ greatly in their resolving power, but for any one system it is the percentage change that matters most.

On most continua Weber's law holds over a substantial portion of the stimulus range (for loud-

ness and brightness over at least 99.9 per cent of the range), but the law fails near the absolute threshold, where discrimination is relatively gross. Nevertheless, Weber's law stands as one of the oldest and broadest empirical generalizations of psychophysics—psychology's "law of relativity," as one writer put it (S. S. Stevens 1951).

**Fechner's law.** If discrimination has seemed to receive more than reasonable attention among students of the senses, the explanation is likely to be found in the significance that the founder of psychophysics attached to the subject. For Fechner, discrimination provided the key to the measurement of sensory magnitude. He began with the postulate that on a given sensory continuum all just-noticeable differences (*jnd*) represent subjectively equal units. Subjective equality of *jnds* is a powerful (if questionable) assumption, because the integration of such units would provide a true scale of subjective magnitude. Since by Weber's law a *jnd* corresponds to a constant fractional increase in the stimulus, it follows that the number of *jnds* grows in an arithmetic series when the stimulus intensity grows in a geometric series. Fechner concluded, therefore, that the magnitude of sensation is a logarithmic function of the stimulus. The logarithmic function implies that equal *ratios* of stimulus magnitude give rise to equal *differences* in subjective magnitude.

**Plateau's power function.** In contrast to the indirectness of the Fechnerian approach to the measurement of sensory magnitudes was an early experiment by the Belgian physicist Joseph A. F. Plateau, who asked a group of artists each to paint a gray that seemed to lie midway between a white sample and a gray sample (a version of a scaling method later termed *equisection*). Of historical interest is Plateau's conclusion that sensation grows as a power function rather than a logarithmic function of the stimulus. But the power function was subsequently given up by Plateau and virtually forgotten until the 1950s. With new techniques for the direct assessment of sensory magnitude, it was shown that Plateau's early conjecture about the form of the psychophysical function happened to be correct. The current approach to the problem is generally to regard as separate properties the resolving capacity of the sensory system and the functional dependence of sensory magnitude on the stimulus.

### The static invariances of sensory systems

A third major problem of psychophysics is to determine those arrangements of stimuli that produce responses that are equivalent in some respect.



The goal of this kind of measurement is to specify all the energy configurations in the environment that produce an invariant or equivalent sensory response.

For example, the goal may be to determine the combinations of intensity and duration of a flash target that produce the same apparent brightness. The level or the duration of a comparison flash is adjusted to match the brightness of a standard flash of fixed intensity and duration. The judgment requires a degree of abstraction because the task is to match for brightness without regard to a difference in apparent duration. A plot relating the duration and intensity that produce a constant (standard) brightness provides an example of an *equal sensation function*. Usually it is desirable to map the family of these equal sensation functions for a pair of parameters. In the present example this means that a function is obtained for each of a set of representative standard brightnesses along the brightness continuum. We learn from this family that, up to a critical duration (roughly 150 milliseconds), a decrease in the stimulus level can be offset by lengthening the flash. Moreover, the critical duration gets systematically shorter as the brightness is increased.

Measurement of the static invariances is common in psychophysics. Examples include the equal brightness functions relating energy and wavelength, the equal loudness functions relating sound pressure and tonal frequency, the equal pitch functions relating frequency and sound pressure (within limits, the apparent pitch of a tone can be altered by a change in sound pressure level), and the equal hue functions relating wavelength and light intensity. (The change in hue when intensity is altered has long been known as the Bezold-Brücke phenomenon.)

The measurement of invariance may call for complete equivalence. An example is the concept of metamerism in color vision. The measurement of metameric pairs (sample lights of identical appearance but different wavelength compositions) has made it possible to state the laws of color mixtures and to predict the color of a sample of any spectral composition.

**Methods of measurement.** Because of its speed and immediacy, the method of adjustment usually recommends itself for the mapping of equivalents. Other usable procedures, however, include constant stimuli, limits, and tracking. The measurement of equivalence is straightforward in principle, but any procedure is usually beset by constant errors, such as the "time error" (see Figure 1).

## Dynamic properties of sensory systems

It is one thing to know the stimulus conditions that produce an invariant sensory effect and another thing to know how much larger one sensory effect is than another—e.g., how much brighter one luminance level appears than another or how much two tones seem to differ in pitch. A major problem of psychophysics is to learn how much the magnitude of the sensory response grows when the stimulus intensity increases.

**The direct scaling methods.** In the 1930s it became apparent to students of hearing that the logarithmic function fails to agree with the reports of observers who are asked to judge the relative loudness of stimuli. Attempts were made to measure the loudness function by a variety of direct methods. Subsequently, the direct methods were expanded and refined and finally applied to the study of all the major sensory continua.

The main feature of the direct methods is the attempt to match segments of the number continuum directly to segments of the sensory continuum. In the method of *magnitude estimation*, various fixed levels of the stimulus are presented one by one in irregular order, and the observer attempts to assign numbers to these levels in proportion to their subjective magnitude. The inverse of this procedure is *magnitude production*: a set of numbers is called out one by one to the observer, who adjusts the level of the stimulus so as to produce subjective magnitudes that are proportional to the numbers. In *ratio production*, a comparison stimulus is adjusted to appear in some fractional or multiplicative relation to a standard stimulus, and in *ratio estimation*, the observer estimates numerically the apparent ratio that corresponds to a pair of stimulus magnitudes. Variations on these procedures are numerous (S. S. Stevens 1958).

**Two classes of continua.** For a few continua, of which pitch is a noteworthy example, a scale of integrated *jnds* turns out to agree well with direct judgment. S. S. Stevens (1957) called these continua *metathetic* and distinguished them from the large class of *prothetic* continua on all of which the *jnd* does not afford a constant unit of subjective magnitude. (Table 1 provides a partial list of prothetic continua.)

**The psychophysical power functions.** S. S. Stevens has also proposed a general psychophysical relation pertaining to all prothetic continua (1957). Equal stimulus ratios are held to correspond to equal sensation *ratios* (rather than to equal sensation *differences*, as Fechner had conjectured). In

other words, the apparent magnitude  $\psi$  grows as a power function of the stimulus magnitude, or  $\psi = k\phi^\beta$ , where  $k$  is a constant of proportionality and  $\beta$  is the exponent. The size of  $\beta$  varies from one continuum to another. In Figure 2A are plotted the power functions in linear coordinates for three continua: apparent length ( $\beta = 1$ ), brightness ( $\beta = 0.33$ ), and the apparent intensity of an electric current passed through the fingers ( $\beta = 3.5$ ). When plotted in log-log coordinates, as in Figure 2B, these same functions become straight lines whose slopes equal the values of the exponents. This is true because the logarithmic form of the power function is  $\log \psi = \log k + \beta \log \phi$ .

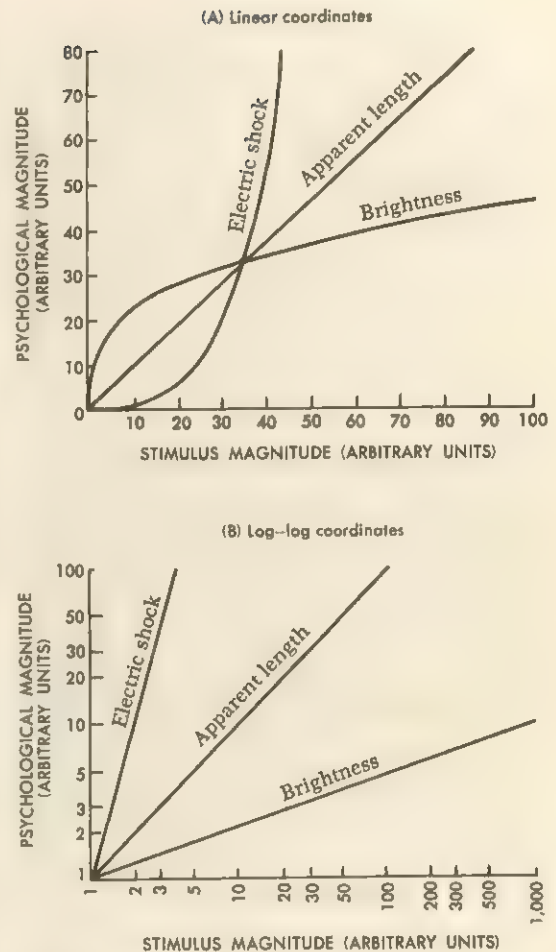
Table 1 shows that the size of the exponent may depend not only on the sense organ stimulated but also on the conditions of the stimulation. Note the difference between the monaural and the binaural loudness functions and the exponent's dependence on frequency for vibration magnitude.

Although the simple equation  $\psi = k\phi^\beta$  holds for large stimulus values, in the neighborhood of the absolute threshold a more precise form is needed. The power equation can be written as  $\psi = k(\phi - \phi_0)^\beta$ , where  $\phi_0$  approximates the absolute threshold. The

**Table 1 — Representative exponents of the power functions relating psychological magnitude to stimulus magnitude on prothetic continua**

Continuum	Exponent	Stimulus condition
Loudness	0.6	binaural
Loudness	0.54	monaural
Brightness	0.33	5° target—dark-adapted eye
Brightness	0.5	point source—dark-adapted eye
Lightness	1.2	reflectance of gray papers
Smell	0.55	coffee odor
Smell	0.6	heptane
Taste	0.8	saccharine
Taste	1.3	sucrose
Taste	1.3	salt
Temperature	1.0	cold—on arm
Temperature	1.5	warmth—on arm
Vibration	0.95	60 cps—on finger
Vibration	0.6	250 cps—on finger
Duration	1.1	white-noise stimulus
Repetition rate	1.0	light, sound, touch, and shocks
Finger span	1.3	thickness of wood blocks
Pressure on palm	1.1	static force on skin
Heaviness	1.45	lifted weights
Force of handgrip	1.7	precision hand dynamometer
Vocal effort	1.1	sound pressure of vocalization
Electric shock	3.5	60 cps—through fingers
Tactile roughness	1.5	felt diameter of emery grits
Tactile hardness	0.8	rubber squeezed between fingers
Viscosity	0.5	stirring silicone fluids
Visual velocity	1.2	moving spot of light
Visual length	1.0	projected line of light
Visual area	0.7	projected square of light

Source: Adapted from S. S. Stevens 1957.



**Figure 2 — The power function for three prothetic continua plotted in (A) linear coordinates and (B) log-log coordinates**

Source: S. S. Stevens 1957.

correction for threshold brings into coincidence the zero of the stimulus scale and the zero of the sensation scale (Luce et al. 1963).

The properties of sensory systems may reveal themselves as parametric shifts in the values of  $\phi$ ,  $k$ , and  $\beta$ . The changes in visual sensitivity that occur under light adaptation provide an example (Stevens & Stevens 1963). Light adaptation causes (1) an elevation in the absolute threshold (i.e.,  $\phi_0$  increases), (2) an increase in the luminance necessary to produce a given subjective brightness (i.e.,  $k$  decreases), and (3) a slight increase in the exponent  $\beta$ . The mapping of these parametric changes has made it possible to write the power



function that pertains to any given level of adaptation and consequently to predict the subjective brightness produced by any luminance level when viewed by an eye adapted to any other luminance level.

**Cross-modality validations.** A method has been devised that circumvents the need for the observer to make numerical estimates of his sensation but leads to the same psychophysical power function (see S. S. Stevens 1961; and Luce et al. 1963). In *cross-modality matching* the task is to make the sensations in two different sense modalities appear equal in strength. The pairs of physical intensities that produce equal apparent intensities can be plotted as an equal sensation function. It turns out that the equal sensation function relating any two prothetic continua  $a$  and  $b$  is itself a power function of the form  $\phi_a = k\phi_b^\gamma$ , where  $\phi_a$  and  $\phi_b$  stand for physical intensity. The size of the exponent  $\gamma$  depends on which two continua are matched. Within the experimental error,  $\gamma$  is predictable from the psychophysical function governing the two continua. Given that  $\psi_a = \phi_a^\alpha$  and  $\psi_b = \phi_b^\beta$  (with suitable units of measurement), and given that  $\psi_a = \psi_b$ , the equation of the equal sensation function becomes  $\phi_a = \phi_b^{\beta/\alpha}$ . The exponent  $\gamma$  thus turns out to be the ratio of the exponents  $\alpha$  and  $\beta$ .

Any continuum could be substituted for the number continuum and used as a "yardstick" to measure sensory magnitudes on all of the other sensory continua. In one set of experiments, for example, force of handgrip as registered on a dynamometer was used to assess subjective magnitudes on nine other prothetic continua (S. S. Stevens 1961). Many other examples could be cited to show that the psychophysical power law is able to predict both the form and the exponent of the equal sensation function obtained by cross-modality matching.

JOSEPH C. STEVENS

[See also SCALING. Other relevant material may be found in ATTENTION; HEARING; INFORMATION THEORY; PAIN; SENSES; SKIN SENSES AND KINESTHESIS; TASTE AND SMELL; VISION; and in the biographies of FECHNER; HELMHOLTZ; TITCHENER; WEBER, ERNST HEINRICH; WUNDT. Statistical techniques applicable to psychophysical methods are described in QUANTAL RESPONSE.]

#### BIBLIOGRAPHY

- BORING, EDWIN G. 1942 *Sensation and Perception in the History of Experimental Psychology*. New York: Appleton.
- FECHNER, GUSTAV T. (1860) 1907 *Elemente der Psychophysik*. 3d ed. 2 vols. Leipzig: Breitkopf & Härtel.
- LUCE, R. DUNCAN; BUSH, ROBERT R.; and GALANTER, EUGENE (editors) 1963 *Handbook of Mathematical Psychology*. Volume 1. New York: Wiley.
- PIÉRON, HENRI (1945) 1952 *The Sensations: Their Functions, Processes, and Mechanisms*. New Haven: Yale Univ. Press; London: Müller. → First published as *Aux sources de la connaissance: La sensation, guide de vie*.
- STEVENS, J. C.; and STEVENS, S. S. 1963 Brightness Function: Effects of Adaptation. *Journal of the Optical Society of America* 53:375-385.
- STEVENS, S. S. (editor) 1951 *Handbook of Experimental Psychology*. New York: Wiley.
- STEVENS, S. S. 1957 On the Psychophysical Law. *Psychological Review* 64:153-181.
- STEVENS, S. S. 1958 Problems and Methods of Psychophysics. *Psychological Bulletin* 55:177-196.
- STEVENS, S. S. 1961 To Honor Fechner and Repeal His Law. *Science* 133:80-86.
- STEVENS, S. S. 1966 A Metric for the Social Consensus. *Science* 151:530-541.
- STEVENS, S. S.; and GALANTER, EUGENE 1957 Ratio Scales and Category Scales for a Dozen Perceptual Continua. *Journal of Experimental Psychology* 54:377-411.
- SWETS, JOHN A. (editor) 1964 *Signal Detection and Recognition by Human Observers: Contemporary Readings*. New York: Wiley.
- SYMPOSIUM ON PRINCIPLES OF SENSORY COMMUNICATION, ENDICOTT HOUSE, 1959 1961 *Sensory Communication: Contributions*. Cambridge, Mass.: M.I.T. Press.
- TITCHENER, EDWARD B. 1905 *Experimental Psychology: A Manual of Laboratory Practice*. Volume 2: Quantitative Experiments. London and New York: Macmillan.
- URBAN, FRIEDRICH M. 1908 *The Application of Statistical Methods to the Problems of Psychophysics*. Philadelphia: Psychological Clinic Press.
- VON BÉKÉSY, GEORG (1928-1958) 1960 *Experiments in Hearing*. New York: McGraw-Hill.

## PSYCHOSIS

It is a generally accepted statistic that about half of all the hospital beds in the United States are occupied by psychiatric patients. Since neurotics are rarely hospitalized, one can define the public health aspect of the problem of the psychoses simply by saying that psychotics occupy about half the beds of the nation on any given day. Approximately half of these beds again are occupied by people diagnosed as suffering from schizophrenia, the major so-called functional psychosis. The majority of the remaining psychiatric beds are occupied by those with a variety of organic psychoses, predominantly aged people suffering from senile or arteriosclerotic psychoses. From the standpoint of the numerical significance of the problem alone, psychoses merit major interest. If one keeps in mind that schizophrenics are likely to become incapacitated early in life, with the mean onset occurring between 20 and 30 years of age, the

problem gains in importance. Although the average length of hospitalization of psychiatric patients decreased by 10 per cent between 1954 and 1964, the number of admissions per year nearly doubled during that period.

Statistics, of course, are only a small part of the problem. They do not tell about the immense human suffering involved in psychoses, nor do they picture the fascination which the topic has aroused from time immemorial. By and large, the term "psychosis" is used as an equivalent of "a major psychiatric disorder," with all the limitations of any such generalization. It has been equated with being possessed by spirits or by the devil, as well as, on the other hand, being revered as a manifestation of divinity. It is also often referred to as "craziness" and is considered by many as a profound revelation of all that goes on in the deep recesses of the human mind.

The fact is that most American clinicians today do not consider a person suffering from a psychosis (a psychotic) to be basically or qualitatively different from so-called normal people. These psychotics may be quantitatively different. They presumably are less able to keep the "deeper" layers of their minds (the primary process) from showing in their daily activities; in some ways, then, they may be more profoundly human.

If one would make a tentative and very limited attempt at defining psychosis, one would have to say it refers to a major psychiatric disorder, one aspect of which is usually some obvious conflict with reality, or a difference from the way a majority of the population perceives reality, or a difference from the expected cultural norms concerning behavior. In the first place, this definition alone makes quite clear that there are areas for disagreement. Not everyone showing such difference or conflict is psychotic. Furthermore, not all people considered psychotic by specialists are necessarily manifestly disturbed in the ways mentioned—either in those actions which are to their own detriment or in those detrimental to society. During most of their lives they may show no obvious sign of any major disorder at all. It may either erupt suddenly or make itself felt only latently and subtly, or be disguised in culturally acceptable forms, such as religious or political fanaticism.

The fact that abnormal behavior of psychotics involves social and cultural norms makes it quite obvious that psychoses are an appropriate concern of such social sciences as sociology, cultural anthropology, and social psychology, and also of allied fields such as jurisprudence and criminology. What is more, psychotic *manifestations* are also

clearly influenced by cultural factors. This holds true not only for different cultures and subcultures but also for temporal changes—for example, within Western culture. What used to be thought possession by the devil has become paranoid ideation involving electric influencing machines, radio, television, radar, or electronic snooping devices.

The *etiology* of psychoses also spans the physical sciences and the social sciences, and there has been much argument over the primacy of organic or "functional" (often equated with psychogenic) causes of psychoses.

The American Psychiatric Association (1952) categorizes psychotic disorders as follows:

*Involuntal psychotic reaction*

*Affective reactions:* manic-depressive reactions (manic type, depressed type, other), psychotic depressive reaction

*Schizophrenic reactions:* simple type, hebephrenic type, catatonic type, paranoid type, acute undifferentiated type, chronic undifferentiated type, schizo-affective type, childhood type, residual type

*Paranoid reactions:* paranoia, paranoid state

## Diagnosis and definition

Psychiatry does not abound in operational definitions, and the term "psychosis," unfortunately, provides no exception. Definition and diagnosis are often closely related in medicine. Therefore, we can say that traditionally and clinically, a patient was diagnosed as psychotic by some simple clinical rules of thumb—if there was a disturbance in orientation regarding time, place, and person; if affect was not appropriate; if the thought processes were found not to be intact (not coherent, relevant, and connected); if the memory of recent or past events was disturbed; if the behavior was inappropriate or obviously influenced by delusions and hallucinations.

This definition includes, of course, both organic and "functional" psychoses. With regard to these terms, there are some more subtle problems of definition, for there are traditionally two basic schools of thought: (1) that psychotics are *qualitatively* different from nonpsychotic people; this concept is held primarily by those who see schizophrenia or other "functional" psychoses as diseases with a clear-cut somatic matrix; (2) that psychotics in general, and schizophrenics in particular, are only quantitatively different from nonpsychotic normal personalities. Adolf Meyer, pioneering the concept of psychobiology, formulated the concept of psychotic reaction types in keeping with the idea that psychoses are a variation of ordinary "reaction types." Most of all, the psychoanalytic



school of thought comprises psychoses within its general theory of personality. Neither school of thought necessarily excludes organic etiologic factors.

Certainly, black-and-white thinking about the concepts "organic" and "functional" seems to be outdated, although the journals hardly reflect this fact. Furthermore, one of the difficulties with the classical scheme of static classification of psychoses is the fact that so many labels are obviously artificial. For, to the chagrin of clinicians, patients often do not seem content to stay in an assigned group, and not only overlap symptomatologically but tend to behave in such a way as to lead clinicians to change their diagnosis of manic-depressive psychosis to schizophrenia, and occasionally vice versa. What is more, diagnoses such as puerperal psychoses, involutional psychoses, arteriosclerotic and senile psychoses often are hardly more than superficial labels. In many patients labeled "psychoses with arteriosclerosis," one is hard put to find evidence of arteriosclerosis, or at least more arteriosclerosis than is found in people of the same age group who are not so diagnosed in vivo or post mortem. In many of these patients, symptoms, character traits, and fragile defenses seem to have been present long before pregnancy, involution, or old age; and the best one can surmise is that the advent of any of these events—hormonal changes, psychological trauma, decreased oxygenation due to an aged circulatory system, decreased social usefulness and esteem—has finally either added the necessary strain or has decreased the integrative capacity, thus producing the change from the latent to the manifest disorder. [See AGING, article on PSYCHOLOGICAL ASPECTS.]

A useful definition—and theory—of psychosis has to accommodate all these observed clinical facts and allow for forces of primarily somatic and primarily experiential factors to play their roles singly and in combination, on different levels of interaction: molecularly as well as grossly (for instance, in the case of bacterial or physical trauma), psychologically as well as sociologically and anthropologically.

In my own thinking, it appears that a *unified theory of psychosis* can be formulated that integrates the organic and the experiential vantage points and offers many advantages.

The conceptual framework, as suggested before, would be that of psychoanalytic ego psychology. The manifestation of any psychosis with any etiology could then be described and understood in terms of changes in ego functioning within a given cultural setting. For this purpose, psychosis is seen

as a *behavioral syndrome*, not a single disease. The somewhat variable symptoms generally associated with this diagnostic label must be understood as the final common path of a number of conditions which may lead to, and manifest themselves in, a severe disturbance of the ego. These conditions may range from a relatively purely psychogenic (experiential) weakness of the ego to afflictions of ego functioning caused by disturbances brought about by infections, by arteriosclerotic, enzymatic, or toxic states, or by traumatic, constitutional, or genetic factors: in short, by any number of chemogenic, histogenic, genogenic, sociogenic, or psychogenic factors, or by any combination thereof. While an outstanding somatic factor may be present, usually this must be accompanied by some psychological predisposition (in terms of ego patterns) to produce the psychotic picture. Conversely, certain somatic factors are probably secondarily brought about by the psychogenic etiology of psychosis, at least for those cases of early onset and severe outcome.

In brief, the multiple-factor, psychosomatic theory of psychoses permits one to understand psychosis as the common result of a variety of individually differing etiologic factors. It enables one to make a prognosis on the basis of a careful study of the specific etiology involved in a specific case, and permits optimal therapeutic control of the highly individual constellation of forces resulting in the manifestly common path of the syndrome, by directly or indirectly, somatically, psychologically, or environmentally, producing better ego functioning.

**Patterns of ego functions and disturbances.** Psychoanalytic theory is predicated upon propositions concerned with the dynamic interaction of environmental and organismic (maturational and congenital) forces with learned responses. Character and personality can thus be understood as patterns of compromise, as resultants of a field of diverse forces. Ego functioning, more specifically, must be described in terms of its pattern: While it may be rather uneven—one ego function may be "better" than another—and the blend ultimately unique, each given pattern may be classified roughly under diagnostic or descriptive headings.

Inasmuch as assessments of the degree of functioning are involved in descriptions of patterns of ego functioning, quantitative assumptions cannot be avoided. Psychoanalytic theory includes many quantitative statements, as might be expected in a theory in which concepts of force and economy play leading roles. The entire concept of libidinal cathexis in object relations and internal representations is, of course, a basically quantitative one.

**Table 1 — Ego functions and their disturbances**

EGO FUNCTIONS	DISTURBANCES
1. Relation to reality <ul style="list-style-type: none"> <li>A. Adaptation to reality               <ul style="list-style-type: none"> <li>a. Differentiation of figure and ground</li> <li>b. Role playing</li> <li>c. Spontaneity and creativeness; regression in the service of the ego</li> </ul> </li> <li>B. Reality testing               <ul style="list-style-type: none"> <li>a. Accuracy of perception</li> <li>b. Soundness of judgment</li> <li>c. Orientation in time, place, person</li> </ul> </li> <li>C. Sense of reality               <ul style="list-style-type: none"> <li>a. Good "self boundaries"</li> <li>b. Unobtrusiveness of ordinary functioning</li> </ul> </li> </ul>	1. Disturbances in relation to reality <ul style="list-style-type: none"> <li>A. Disturbances in adaptive capacity               <ul style="list-style-type: none"> <li>a. Inappropriate behavior with subjective or objective difficulties</li> <li>b. Inability to cope with deviations in normal routine</li> <li>c. Failure in social adaptation; rigidity</li> </ul> </li> <li>B. Disturbance in reality testing               <ul style="list-style-type: none"> <li>a. Projection, rationalization, denial and the distortion of reality by hallucinations and delusions</li> </ul> </li> <li>C. Disturbances in sense of reality               <ul style="list-style-type: none"> <li>a. Feelings of estrangement and lack of spontaneity</li> <li>b. Excessive feelings of déjà vu</li> <li>c. Oneirophrenia</li> <li>d. Cosmic delusions</li> <li>e. Confused body images</li> <li>f. Intrusion of self as subject or object</li> <li>g. Physiological manifestations</li> </ul> </li> </ul>
2. Regulation and control of drives <ul style="list-style-type: none"> <li>a. Ability to engage in detour behavior</li> <li>b. Frustration tolerance (neutralization of drive energy)</li> <li>c. Anxiety tolerance</li> <li>d. Integrated motility</li> <li>e. Tolerance of ambiguity</li> <li>f. Sublimation</li> </ul>	2. Disturbances in drive control <ul style="list-style-type: none"> <li>a. Conduct and habit disorders (temper tantrums, nail-biting, etc.)</li> <li>b. Accident proneness</li> <li>c. Excessive impulsivity</li> <li>d. Tension states</li> <li>e. Catatonic and manic excitement</li> <li>f. Psychomotor slow-up of catatonia and depression</li> <li>g. Lack of or incomplete acquisition of control of excretory functions</li> <li>h. Physiological manifestations</li> </ul>
3. Object relations <ul style="list-style-type: none"> <li>a. Capacity to form satisfactory object relations</li> <li>b. Object constancy</li> </ul>	3. Disturbances in object relations <ul style="list-style-type: none"> <li>a. Psychotoxic and psychic deficiency diseases (in infancy)</li> <li>b. Narcissism, autism</li> <li>c. Symbiotic relationships</li> <li>d. Anaclitic relationships</li> <li>e. Hypercathexis of the self; ambivalence, fear of incorporation, sado-masochism</li> </ul>
4. Thought processes <ul style="list-style-type: none"> <li>a. Selective scanning</li> <li>b. Ability to avoid contamination by inappropriate material or drives</li> <li>c. Good memory</li> <li>d. Sustained ability to concentrate</li> <li>e. Abstracting ability</li> </ul>	4. Disturbances in thought processes <ul style="list-style-type: none"> <li>a. Thinking organized and compelled by drives</li> <li>b. Preoccupation with instinctual aims</li> <li>c. Autistic logic</li> <li>d. Loose and "nonsensical" types of associative links</li> <li>e. Distortion of reality</li> <li>f. Lack of referents in time and place, anthropomorphism, concretism, symbolism, syncretism, etc.</li> <li>g. Magic thinking</li> </ul>
5. Defensive functions <ul style="list-style-type: none"> <li>a. Repression (as a barrier against external and internal stimuli)</li> <li>b. Sublimation, reaction formation</li> <li>c. Projection</li> <li>d. Denial, withdrawal, and other defenses</li> </ul>	5. Disturbances in defensive functions <ul style="list-style-type: none"> <li>a. Emergence of primary thought process</li> <li>b. Overreaction to stimuli</li> <li>c. Déjà vu experiences</li> <li>d. Lack of drive control</li> <li>e. Frightening hypnagogic phenomena</li> <li>f. Increase in parapraxes</li> <li>g. Impairment in emotional control</li> </ul>
6. Autonomous functions <ul style="list-style-type: none"> <li>a. Perception</li> <li>b. Intention</li> <li>c. Intelligence</li> <li>d. Thinking</li> <li>e. Language</li> <li>f. Productivity</li> <li>g. Motor development</li> </ul>	6. Disturbance in autonomous functions <ul style="list-style-type: none"> <li>a. Corresponding impairment of these ego functions</li> </ul>
7. Synthetic functions <ul style="list-style-type: none"> <li>a. To unite, organize, bind, and create—the ego's ability to form gestalten</li> <li>b. Neutralization</li> <li>c. Sublimation</li> <li>d. Somatic "homeostasis"</li> </ul>	7. Disturbances in synthetic functions <ul style="list-style-type: none"> <li>a. Tendency to dissociation</li> <li>b. Inability to tolerate change or trauma</li> <li>c. Inability to "bind" psychic energy</li> </ul>



However, psychoanalysts generally have not addressed themselves to attempts at precise quantification. It should be possible, though, to adapt the experimental methods of measurement to psychoanalytic variables.

Meaningful methods of measuring ego strength are possible if ego strength is defined as "the totality of the ego's capacity to perform its many functions." Each individual ego function could be quantitatively assayed, given a weighted score, and combined into a total score, much as the IQ is arrived at in the Wechsler-Bellevue Intelligence Test. Also, ego strength must be viewed globally, very much like Wechsler's view of intelligence (1939). The ego cannot be conceptualized as a perfect sphere, of course, with each area of ego functioning constituting a radius, nor can it even be compared in this context to the segments of an orange. The image that suggests itself is that of an uneven raspberry on which each surface point constitutes the terminus of one of the many ego functions. Furthermore, this protean raspberry might be made of stretchable rubber which would change its shape developmentally and would be subject to momentary and daily variations.

Table 1 presents some important groups of ego functions and disturbances.

**Multiple-factor theory.** The concept of psychoses as disorders that have many different etiologies but share a final common path of ego disturbance leads quite logically to the concept that the diagnosis of psychosis at present can best be made on the basis of the degree of ego disturbance in a given patient. It also follows that the degree of ego disturbance that we are willing to call psychosis is based on rather arbitrary decisions.

As mentioned, there are variations of ego functioning in every person—that is, healthy people exhibit certain ego functions which are of a very high quality and others which are less so. It is probably true that a surprising number of ego disturbances, sometimes of considerable severity, appear in a large number of people who are generally considered normal statistically and in terms of lifelong functioning. These are still "normal" people, however, by virtue of the fact that the ego disturbance occurs in a relatively small segment of their personality or one which, in their particular setting, does not crucially interfere with functioning. A fear of deep water, for instance, is not likely to be a vital disturbance in an inland state such as Kansas. Obviously, forms and expectations of ego functions will vary culturally.

Also, one may find, on taking a careful life history, that there was some episode in ego disturb-

ance at a certain age which, although fairly severe, was self-limited and had no further consequence. Chance is involved in such matters: a marked disturbance in military service may become a matter of troubling public record and later secondary pathology; some private disturbance may forever remain unknown and become a subject of amnesia.

Under certain conditions, e.g., those of extreme deprivation, some delusions or hallucinations (of food, water, companions) must be considered virtually as adaptive functions of the ego which usually promptly disappear when the emergency does; they may in fact have definite survival value during the emergency: for example, they may bolster the person with the irrational feeling that a special deity is guarding him and with the certainty that all will be well. Then there are, of course, a wide variety of neurotic ego disturbances that must be considered. All these facts will serve to highlight further the difficulties in forming a diagnosis of psychosis on the basis of one specific syndrome.

It is to be remembered that from a theoretical viewpoint we postulate that psychosis is not one condition, not one point on a continuum of ego strength, but a *range* along a continuum: within a group diagnosed as psychotic one could classify those who are sicker and less sick, patients with more and less over-all ego strength (also varying ego strength in different areas at different times). In this sense ego functioning not only must be seen as a quantitative proposition but also must be viewed as greater or lesser in different areas *at different times*.

It is useful to remember that, in essence, the diagnosis of psychosis is a phenomenological diagnosis predicated upon the observable failure of ego functions (including reported ones, of course), rather than upon assumptions concerning dynamic structure or other criteria.

**The diagnostic continuum.** Psychosis can be conceptualized as ranging at the low end of the continuum of ego strength, while normality covers a range at the other end. Presumably, schizophrenia, manic-depressive psychosis, obsessive-compulsive neurosis, severe character disorders, phobias, anxiety hysterias and neuroses, and hysterias lie approximately in that order, reading from left to right, between these two points. Figure 1 depicts this continuum. However, this traditional viewpoint is of limited value, since each of these conditions is characterized only by a symptom picture that expresses a single outstanding defensive compromise formation. There can be no question that a severe hysteria may involve a weaker ego and may

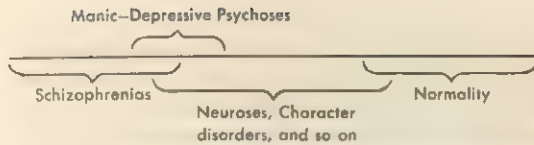


Figure 1 — A schematic representation of the diagnostic continuum (oversimplified and with no quantitative implications)

be a condition relatively closer to a psychosis (schizophrenia) than to a mild obsessive-compulsive disorder or a circumscribed phobia.

A psychosis may sometimes be difficult to diagnose in its early stages because the clinical picture may involve a vast variety of patterns of ego functioning. These might, of course, also be present in neurosis, circumscribed neurological disorders, character disorders, psychopathic (sociopathic) personalities, perversions, and other disorders, and the psychosis may overlap with them.

It is not the intention here to discuss in detail the clinical problems of differential diagnosis but, rather, to attempt a conceptual clarification: the clinical diagnosis of early psychosis must be predicated upon the judgment that the degree of ego disturbance is so severe and the defenses so pathological that one may expect the development of a full-blown psychosis if there is no skilled intervention.

Just as there are many varieties of full-blown psychoses, so there are many forms of early psychotic symptomatology. Sociopathic irresponsibility is indicative of poor integration of the superego and poor frustration tolerance; poor reality testing and poor object relations (by the ego) may occur in a person previously conscientious, polite, and kind to children and animals. Thus tabloids frequently print lurid accounts of choir-member offspring of pillars of the community who, apparently, suddenly become psychotic and commit criminal acts. Yet it is quite possible that signs of disturbance had been present for a long time. Seclusiveness, as well as sudden extreme "extroversion," may be a first indication of a psychotic break. Lability of mood often signifies impaired control.

Nor are the specific manifestations of ego disturbance the only prodromal signs of psychosis: the prepsychotic personality may be sociopathic, infantile (little control, many pregenital drives), brilliantly high-strung (artistic or intellectual achievement at the cost of other development), excessively religious (i.e., a reaction formation to poorly controlled hostility), dull and apathetic (because of early disturbance of autonomous functions

and object relations), or have poor achievement (because of preoccupation with fantasies, etc.), despite high intelligence.

A good account of the great variety of full-blown psychoses has yet to be written. In essence, one psychotic may differ almost as much from another as any individual may differ from another. There are those who never progressed, and those who regressed from the highest levels of achievement; those who are obviously bizarre and hard to empathize with, and those whom one is likely to consider less disturbed than they actually are because of a good façade. Some psychotics are very kind; others are cold and sadistic. Some are very productive and sensitive; others are seemingly without drive or ability. Psychotics have been unjustly downgraded and romantically overestimated as artists, intellectuals, or, as some would have it, revolutionaries whom society of necessity secludes. It is easier to be categorical about psychotics if one does not have too much understanding and knowledge of them.

**Differential diagnostic concepts.** In the diagnosis of psychosis a number of diagnostic labels of secondary order are frequently used, sometimes in unclear and confusing ways.

**Borderline psychosis.** Borderline psychosis can be quite accurately defined as a condition of ego strength and personality functioning which lies on the continuum between a psychotic and a neurotic condition. In an appraisal of the ego strength of such a patient, one finds that thinking, affective control, perceptual reliability, and object relations are all of questionable soundness without being quite disturbed enough to merit the diagnosis of psychosis.

The diagnosis of borderline psychosis is indicated if the condition described refers to an apparently stable condition of functioning, perhaps even a lifelong one, without signs of progressive deterioration. In that sense, "borderline" is identical with the term *psychotic character*, and it relates to a character formation with many unfortunate defensive traits that are likely to make a person appear quite peculiar (even though he may often be quite brilliant and useful in isolated areas) without his ever developing a frank psychosis.

**Incipient psychosis.** If the picture of dubious functioning described as borderline psychosis is present and is accompanied by signs of increasing lability of mood, progressive inability to control impulses, increasing emergence of the primary process, and increasing crumbling of defenses, with *déjà vu* experiences, feelings of unreality, impaired sleep, impaired appetite, and a rising anxiety level,



then we are dealing with incipient psychosis, and steps must be taken to avoid its further progress.

This discussion implies quite clearly that in diagnosis it is important to judge not only the nature of functioning and the quality of the defense patterns but also the *stability* of the defensive patterns. Clinical experience permits little doubt that many people arrive at certain character formations—more or less pathological or normal—which they maintain all their lives. Not only do some people remain “borderline” all their lives, but some psychotics remain stabilized at certain levels of illness and do not regress further; others do regress. Some psychotics have four, five, or more episodes during their lives which always take the same form and which then show spontaneous remission. And some people have character peculiarities—psychotic or otherwise—which are lifelong, without any progression or regression of note. Clinical experience suggests the hypothesis that each person has a certain amount of ego strength—or, more specifically, strength of the synthetic functions of the ego—and this very often involves the stabilizing at a certain level of pathology, below which a person does not go under ordinary or possibly even extraordinary life circumstances.

*Potential psychosis.* Potential psychosis seems a useful diagnosis if one can observe poor ego functioning together with defenses that could eventually lead to a psychosis. That is, one might note the same symptoms that occur in borderline conditions, but accompanied by less stability in the defensive pattern. In the potential psychotic one may note previous episodes of increasing precariousness of balance between drive and defense, with outstanding defenses of denial and projection, and signs of loss of perceptual and motor control. In other words, the diagnosis “potential psychosis” may be useful to connote a condition ranging between “borderline” and “incipient,” that is, more labile than the former and less labile than the latter.

*Latent psychosis.* The designation “latent psychosis” might be employed when there is evidence of an existing psychosis which is, for practical purposes, covered up most of the time: i.e., in psychotherapy one might see that somebody is usually under the influence of primary process thinking and constantly beset by severe distortions of reality, but he manages to keep all these problems private by the use of intelligence and clever rationalizations. Such a person requires a favorable environment in order to continue to appear manifestly normal. Usually a forced change in environment or a change in living circumstances will suddenly disrupt the delicate arrangement; the pa-

tient will then surprise the world with the emergence of full-blown delusions of obviously long standing. The concept of latency thus refers in essence to the social impression which this individual makes. This nonpsychotic impression may also prevail in an ordinary clinical examination if the patient feels uncooperative, but it is extremely likely to appear in projective and other diagnostic testing, and in prolonged psychoanalytic interviews.

### Therapy and prevention

The problem of a systematic and effective therapy for the psychoses is, of course, related to the matter of etiology: one usually wants to affect the causes of a disturbance. Since there is no general agreement on the causes of the “functional” psychoses and even very little on the etiology or pathogenesis of the organic psychoses, e.g., the so-called arteriosclerotic or senile psychoses, there is not much consensus about therapy for any of them.

If one accepts the proposition that there are multiple etiologies, it follows that one must attempt to treat each patient as much as possible within the individual constellation of causative factors.

A few therapies enjoy particularly widespread usage. Of the “shock treatments,” insulin therapy has lost the importance it had in the decade after World War II, but it is still used in some cases of schizophrenia. Its relative expense in terms of personnel and the need for considerable experience essential for its use are responsible in part for its decline. Electroshock therapy, or, as it is more often called, ECT (for electroconvulsive therapy), at first replaced insulin treatment to a large extent both for schizophrenia and, especially, for depressions. In turn, it lost popularity with the development of psychotropic drugs in the 1950s. By 1965 the tide seemed to have turned to a certain extent. The anti-depressants, for instance, have not proved miracle drugs after all, although undoubtedly they are often clinically useful (Cole 1964). ECT has been greatly refined by use of intravenous anesthetics and muscle-relaxing drugs, obviating the panic some patients felt and the danger of fractures. [See ELECTROCONVULSIVE SHOCK; MENTAL DISORDERS, TREATMENT OF, *article on SOMATIC TREATMENT.*]

The psychotropic drugs were the most important therapeutic modality in the mid-1960s. Popularly, the “tranquilizers” are most widely known and, among them, especially the phenothiazines. There is little question that the psychotropic drugs have changed much of psychiatric practice, for they often decrease or remove psychotic symptoms, make disturbed patients behave without violence, and decrease panic, delusions, and hallucinations.

Thus, state hospital stays have been decreased, rehabilitation is markedly helped, and hospitalization in general hospitals is made possible.

Nevertheless, the psychotropic drugs are by no means curative, nor is their over-all effectiveness proved beyond doubt or the mode of their beneficial effect agreed upon. Empirically, however, they seem to create at least the conditions in which other therapy, spontaneous improvement, or at least socially acceptable behavior is possible for many psychotics who could not profit similarly without them.

As in all therapy, controlling variables other than the therapeutic agents, providing comparable groups of patients, and ruling out the effect of time alone are very difficult. The most promising development might well be Chassan's development of the intensive design (Chassan & Bellak 1966). [See DRUGS, *article on PSYCHOPHARMACOLOGY*.]

Social aspects of therapy have found their expression especially in community psychiatry (Bellak 1964). Keeping the patient close to home and providing early treatment in community clinics and follow-up in the community after a rather prompt hospital discharge play a marked role. Rehabilitation programs, as well as group therapy and family-centered therapy, play an often-constructive role.

The psychotherapy of psychoses plays more of a role for the extramural patient, if for no other reason than that the large state hospitals still rarely have enough staff for it. In clinic and private office practice, however, it often leads to empirically satisfying results. The adjunct treatment, by drugs, milieu therapy, and other measures, is becoming more popular.

For some of the organic psychoses, such as acute alcohol psychoses, there are specific, somatic treatment forms.

It is difficult to generalize about the effectiveness of treatment of psychoses as a group. In the case of an acute schizophrenic or depressive psychosis, it is not unusual today that as high as 80 per cent of all first admissions to a therapeutically active hospital may be discharged within 30 days. On the other hand, for about 7 per cent of all schizophrenics, the length of hospitalization still remains 20 years.

Primary prevention in the public health sense is also closely related to concepts of etiology. Since even most geneticists admit the significance of environmental factors at least for the manifestation of latent liabilities, and hardly anybody doubts the effect of experiential factors, a great deal more could be done for prevention of psychoses by public education, psychiatric well-baby clinics, and, where

need be, legislation concerning the disposition of children with unsuitable parents, or whose parents are divorced or otherwise separated.

Secondary prevention, by early treatment of incipient disturbances, is making progress by virtue of greater federal and state investment in community mental health. Similarly, tertiary prevention (of chronicity) by rehabilitation is in the process of development.

### Social and legal aspects

As long as psychotics were considered to suffer from a disease, they were usually studied as individuals in isolation, as it were. With an increasing awareness of the importance of interpersonal relationships, both the genesis and the cause of psychosis in relation to social interaction have attained more importance. Also, the ward setting and its effect on psychoses have been described by Stanton and Schwartz (1954), and social and epidemiological aspects of psychoses have been formulated by Hollingshead and Redlich (1954), Faris and Dunham (1939), Lemkau and Crocetti (1958), and others. The differing patterns in the family structure, especially of schizophrenics, have been discussed, e.g., by Opler and Singer (1956) and Sanua (1963). [See PSYCHIATRY, *article on SOCIAL PSYCHIATRY*.]

Because of the impairment of judgment and the lack of impulse control in many psychotics, and the occasional commission of crime by psychotics, special social-legal problems exist with regard to the involuntary commitment of the mentally ill to hospitals and the observation of their constitutional rights under such circumstances. Furthermore, the specific problem of legal responsibility for crimes has to be considered [Roche 1964; *see also PSYCHIATRY, article on FORENSIC PSYCHIATRY*].

### Research

Research in psychoses has suffered greatly from a lack of conceptual clarity. And a lack of methodological sophistication among clinicians who usually engaged in research only as a by-product of clinical and therapeutic work has bedeviled the field. As examples one might mention the tremendous popularity of Kretschmer's morphological classification (athletic, asthenic, pyknic) and the complete absence of valid supporting data for it from controlled studies. The same can be said for the welter of genetic studies: if anywhere there are diagnoses of family members which have been made under blind, controlled conditions so as to avoid contamination of diagnostic judgment of the incidence of psychoses in the relatives of the pa-



tient (as compared with a control group), I have never run across them. As Tienari points out (1963), most of the genetic studies have been particularly poor methodologically. [See MENTAL DISORDERS, *article on GENETIC ASPECTS; PSYCHOLOGY, article on CONSTITUTIONAL PSYCHOLOGY.*]

There are some methodologically sophisticated studies by social scientists and by physicians with training in the various medical and basic sciences. These researchers do not usually have clinical psychiatric experience and are rarely psychodynamically trained. Rather than cast aspersions on their undertakings, one simply has to acknowledge that psychoses are protean manifestations of all facets of human life and that research has suffered from the fact that it is difficult to marshal all the knowledge and experience needed to deal with the many facets involved.

There is no bodily system or organ which has not been more or less extensively implicated in the cause of psychoses, especially of schizophrenia (Bellak 1948). Psychoanalysts and psychologists have described psychogenic mothers, fathers, and general family settings, aside from specifics of individual development (Lidz et al. 1956). Sociologists have shown relationships to urban and rural populations and socioeconomic classes (Faris & Dunham 1939). Cultural anthropologists differ on whether more advanced cultures may have a greater incidence of psychoses than do primitive ones (Benedict 1958). Semanticists have related schizophrenia to the double bind and to semantic confusion. Geneticists have had a field day only occasionally interrupted by careful studies (Tienari 1963). Physical anthropology, by way of Kretschmer and Sheldon, has not lacked a voice, nor has such a factor as birth date (season of the year) been neglected (Barry & Barry 1964). In fact, hardly a week goes by without some etiological claim being mentioned in the popular or scientific press, and hardly a year without the total crop of claims being thoroughly discredited.

Just as the implication of the spirochetes in general paresis and the ensuing fever treatment raised the hope for similar etiology and therapy of psychoses, so the physical treatment modalities, by inference, were held a promising road for the uncovering of etiology. Sakel, Meduna, Cerletti and their followers were thus influenced, respectively, by insulin, metrazol, and electric shock treatment; they all formulated etiological theories which did not bear the burden of systematic investigation. The latest flurry of this type of research was produced by the advent of psychotropic drugs: if tranquilizers produce symptomatic changes in psychotics, and if these tranquilizers seem to have

an effect, for example, on serotonin metabolism, then one might suggest that a serotonin disorder produces psychoses. By the same reasoning, if monoamine oxidase (MAO)-inhibiting drugs benefit depressions, it was felt that maybe MAO is etiologically implicated in the production of depression. Of course, it quickly turned out that different tranquilizers have their effect by different pathways and that drugs not affecting MAO also could lift depressions (for instance, amitriptyline). [See MENTAL DISORDERS, *article on BIOLOGICAL ASPECTS.*]

Some of the most ambitious research was carried out by means of the so-called model psychoses and experimental analogues. Long ago, it was known that certain drugs, notably mescaline, bulbocapnine, and others could produce psychoses not unlike schizophrenic psychoses. The most widely heralded drug in the 1960s was LSD (lysergic acid diethylamide-25). The most ambitious hope was to produce psychoses experimentally, then to be able to abolish them with a tranquilizer. As it happens, LSD-induced psychosis has only little resemblance to schizophrenia (except in borderline schizophrenics, in whom LSD, like many other traumata, can apparently induce a genuine psychosis, if one is not cautious enough to screen subjects). Other physiological experimental investigations only suffered from oversimplification and artifacts.

To confound the issue further, or, rather, to clarify the fact that psychosis is apparently a condition which is merely the final common path of many different effects on ego functions, interesting experiments in sensory deprivation or perceptual isolation (Lilly 1956) showed that psychosis-like conditions with hallucinations and delusions could be induced in some subjects put into these experimental conditions. [See PERCEPTION, *article on PERCEPTUAL DEPRIVATION.*]

The subject of research in psychosis can, therefore, be summarized by saying that hardly any of a worthwhile nature exists so far, with the exception of some rather careful work refuting a variety of heuristic claims (Kety 1959). Such claims have usually been the result of oversimplification, lack of controls, and lack of sophistication. It is my strong personal belief that progress will come only when an integrated, highly coordinated, multidisciplinary, extremely extensive and intensive approach is made, preferably under one roof, such as a "national institute for psychoses research." Because of a lack of coordination and integration of research, the awarding of individual research grants has so far shown nothing but disappointment, from any basic standpoint.

LEOPOLD BELLAK

[Directly related are the entries DEPRESSIVE DISORDERS; MENTAL DISORDERS; NEUROSIS; SCHIZOPHRENIA. Other relevant material may be found in MENTAL DISORDERS, TREATMENT OF; MENTAL HEALTH; PSYCHIATRY; PSYCHOANALYSIS, especially the article on EGO PSYCHOLOGY; PSYCHOSOMATIC ILLNESS; and in the biographies of FREUD; MEYER; SULLIVAN.]

# BIBLIOGRAPHY

- AMERICAN PSYCHIATRIC ASSOCIATION, COMMITTEE ON NOMENCLATURE AND STATISTICS (1952) 1963 *Diagnostic and Statistical Manual: Mental Disorders*. Washington: The Association.
- BARRY, HERBERT; and BARRY, HERBERT JR. 1964 Season of Birth in Schizophrenics in Relation to Social Class. *Archives of General Psychiatry* 11:385-391.
- BELLAK, LEOPOLD 1948 *Dementia Praecox*. New York: Grune.
- BELLAK, LEOPOLD (editor) 1958 *Schizophrenia: A Review of the Syndrome*. New York: Logos. → Reviews work in the field between 1946 and 1956; contains 4,000 references.
- BELLAK, LEOPOLD (editor) 1964 *Handbook of Community Psychiatry and Community Mental Health*. New York: Grune.
- BENEDICT, PAUL K. 1958 Socio-Cultural Factors in Schizophrenia. Pages 694-729 in Leopold Bellak (editor), *Schizophrenia: A Review of the Syndrome*. New York: Logos.
- CHASSAN, J.; and BELLAK, LEOPOLD 1966 An Introduction to Intensive Design of the Evaluation of Drug Efficacy During Psychotherapy. Pages 478-499 in Louis A. Gottschalk and Arthur H. Auerbach (editors), *Methods of Research in Psychotherapy*. New York: Appleton.
- COLE, JONATHAN 1964 Efficacy of Antidepressant Drugs. *Journal of the American Medical Association* 190: 448-455.
- PARIS, ROBERT E. L.; and DUNHAM, H. WARREN (1939) 1960 *Mental Disorders in Urban Areas: An Ecological Study of Schizophrenia and Other Psychoses*. New York: Hafner.
- HOLLINGSHEAD, A. B.; and REDLICH, F. C. 1954 Schizophrenia and Social Structure. *American Journal of Psychiatry* 110:695-701.
- KETY, SEYMOUR S. 1959 Biochemical Theories of Schizophrenia. *Science* 129:1528-1532, 1590-1596.
- LEMKAU, PAUL V.; and CROCETTI, GUIDO M. 1958 Vital Statistics of Schizophrenia. Pages 64-81 in Leopold Bellak (editor), *Schizophrenia: A Review of the Syndrome*. New York: Logos.
- LIDZ, THEODORE; PARKER, BEULAH; and CORNELISON, ALICE 1956 The Role of the Father in the Family Environment of the Schizophrenic Patient. *American Journal of Psychiatry* 113:126-132.
- LILLY, JOHN C. 1956 Mental Effects of Reduction of Ordinary Levels of Physical Stimuli on Intact, Healthy Persons. *Psychiatric Research Reports* 5:1-9.
- MEYER, ADOLF 1948 *The Commonsense Psychiatry of Dr. Adolf Meyer*. Edited by Alfred Lief. New York: McGraw-Hill. → Consists of 52 selected papers.
- OPLER, MARVIN K.; and SINGER, JEROME L. 1958 Ethnic Differences in Behavior and Psychopathology: Italian and Irish. *International Journal of Social Psychiatry* 2:11-23.
- ROCHE, PHILIP Q. 1964 Psychiatry, Law and Community. Pages 409-430 in Leopold Bellak (editor), *Handbook of Community Psychiatry and Community Mental Health*. New York: Grune.

SANUA, VICTOR D. 1963 The Socio-Cultural Aspects of Schizophrenia: A Comparison of Protestant and Jewish Schizophrenics. *International Journal of Social Psychiatry* 9:27-36.

STANTON, ALFRED H.; and SCHWARTZ, M. S. 1954 *The Mental Hospital: A Study of Institutional Participation in Psychiatric Illness and Treatment*. New York: Basic Books.

TIENARI, PEKKA 1963 Psychiatric Illnesses in Identical Twins. *Acta psychiatrica scandinavica* 39 (Supplement 171). → The entire issue is devoted to Tienari's study.

WECHSLER, DAVID (1939) 1958 *The Measurement and Appraisal of Adult Intelligence*. 4th ed. Baltimore: Williams & Wilkins. → First published as *The Measurement of Adult Intelligence*.

# PSYCHOSOMATIC ILLNESS

Investigations into psychosomatic illness have focused on the relationships between emotional life and bodily processes—both normal and pathological—and not on the isolated problems of the diseased mind or the diseased body. The mind-body dichotomy is eliminated through the thesis that there is no duality of mind and body, mental and physical, but only a unity of the total being. It is assumed that the physiology of mood, instinct, and intellect differs from other physiology in degree of complexity but not in quality. Hence, while divisions of medical disciplines such as physiology, internal medicine, and psychiatry may be convenient for academic administration, biologically and philosophically these divisions have no validity. Psychic and somatic phenomena take place in the same biological system and are two aspects of the same process.

The ever-increasing flow of observations concerning the relation of psychological and physiological processes in the human organism supports this orientation. The influence of specific emotional tensions upon biochemical, endocrinological, and physiological changes has been studied both clinically and experimentally, as has the influence of physiological changes upon the emotional life of man.

Modern endocrinology and biochemistry may well be the progeny of the earlier humoral theory that medicine sought to refute or disregard. The whole man, not just his cells, tissues, and organs, becomes the subject of study for the modern psychosomatic researcher. Facts are seen as elements of the whole. The gestalt of man is studied with the appreciation that the whole is more than the sum of the parts; but these parts, nonetheless, are integrated in such fashion as to contribute to the whole. Personality is the expression of the unity of the organism. Neurology, neuropsychology, general physiology, endocrinology, biochemistry, pharmacology, and genetics contribute to our knowledge



of the structure and function of the body. Psychology, psychiatry, and psychoanalysis contribute knowledge of the subjective phenomena which are reflections of physiological processes but are also the products of past social, cultural, and interpersonal relationships that have become internalized and thus are part of the integrity of the organism.

Franz Alexander observed:

The body, that complicated machine, carries out the most complex and refined motor activities under the influence of such psychological phenomena as ideas and wishes. The most specifically human of all bodily functions, speech, is nothing but the expression of ideas through a refined musical instrument, the vocal apparatus. All our emotions we express through physiological processes; sorrow, by weeping; amusement, by laughter; and shame, by blushing. All emotions are accompanied by physiological changes: fear by palpitation of the heart; anger by increased heart activity, elevation of blood pressure and changes in carbohydrate metabolism; despair by a deep inspiration and expiration called sighing. All these physiological phenomena are the results of complex muscular interaction under the influence of nervous impulses, carried to the expressive muscles of the face and to the diaphragm in laughter, to the lacrimal glands in weeping, to the heart in fear, and to the adrenal glands and to the vascular system in rage. The nervous impulses arise in certain emotional situations which in turn originate from our interaction with other people. The originating psychological situations can only be understood . . . as total responses of the organism to its environment. (1950, pp. 38-39)

Various somatic manifestations may accompany different transitory emotional states in the normal organism. Thus, disturbances of the stomach, bowels, cardiovascular system, or respiratory system may be expressions of anxiety. These upsets are usually reversible, and in such cases morphological changes of the cells, tissues, or organs involved are nonspecific. The anatomical structure is not permanently or grossly altered; only the function is disturbed. Since these functional disturbances are triggered by emotional factors, the psychological understanding of the diseased patient is necessary if one is to assist him fully. This may not require formal psychotherapy, since the removal of the emotionally stressing external situation may by itself re-establish a level of equilibrium. When external rearrangements are not possible, the understanding of the anatomical and physiological mechanisms involved may allow for biochemical and pharmacological interventions that can temporarily interfere with abnormal pathways of discharge and thus alleviate symptoms during the period of equilibrium re-establishment.

Often an emotionally disturbing situation triggers a conflict which exacerbates a previous con-

flict constellation that has been rendered unconscious and largely inactive. In such cases, the functional distress may not be easily reversed, and chronic symptomatology results. Thus, a functional disturbance of long duration or of overwhelming intensity may lead to definite and demonstrable anatomical changes and to the clinical picture of severe organic illness. Research in such diseases as duodenal ulcer, ulcerative colitis, bronchial asthma, neurodermatitis, essential hypertension, rheumatoid arthritis, thyrotoxicosis, *diabetes mellitus*, glaucoma, migraine, and *anorexia nervosa* indicates the probability of particular personality constellations having particular vulnerability to specific conflict situations.

The particular conflicts associated with the individual disorders have been detailed in *Studies in Psychosomatic Medicine* (Alexander et al. 1948). In peptic ulcer, for example, it has been found that the individual has a basic core of infantile oral dependency, which he copes with in various ways, such as overcompensation, aggression, or assertiveness. When such an individual's coping mechanisms are impaired, or his oral-dependent cravings increase, he may no longer manage the internal conflict at a psychological level. Increased gastric secretion and/or gastric vasoconstriction may give rise to symptoms of gastric distress. If there is a decrease in stress on the organism, or a bolstering of the internal or organic system, strain may be diminished without symptomatic progression and with symptom remission. If psychic stress continues, what was originally dysfunction may result in an actual organic lesion, the peptic ulcer. The oral-dependent characteristics can be detected in dreams, behavior, and general character structure.

Other conflicts have been linked with particular diseases: repressed hostility with essential hypertension, fear of separation with bronchial asthma, and fear of destruction of the self with thyrotoxicosis. These conflicts have been described in detail in *Psychosomatic Specificity* (1968).

For the present, the functional theory of organic disorders associated with the psychosomatic concept includes the recognition of three etiological components acting together. Certain *external causative* factors can act as precipitators of imbalance and give rise to functional or structural disruption. These external factors have their effect upon underlying *internal systems* of equilibrium and integration. Thus many chronic disturbances are not caused by external, mechanical, chemical, or infectious agents, but by the stresses of the struggle for existence. These emotional conflicts arise during daily living, in the social interchange with signifi-

cant figures in the environment that have symbolic as well as actual significance. Finally, one must bear in mind the *predispositions* to disequilibrium which reflect constitutional and experiential influences upon the organism. These include genetic factors, the influence of intrauterine and extrauterine physiological and pathological agents, the crucial parent-sibling relationships during critical personality developmental periods of the first decade of life, especially as they relate to mental functioning, and the impact of personal, social, and cultural influences in child-care and child-rearing practices. These predisposing factors constitute the underlying matrix of vulnerability to certain situations occurring later in life. Continuous and repressed fears, aggressions, and libidinal wishes can, for susceptible individuals, result in permanent chronic emotional tensions which disturb the functions of various organ systems. In this way such activities as digestion, respiration, and circulation may at first show signs and evoke symptoms of functional distress; later, if nothing is done to alleviate the disturbed situation, organic disease may follow.

Many researchers are concerned with the study of specific emotional conflicts and personality organizations as they relate to particular organic syndromes. For example, the cardiovascular responses intimately linked with rage seem to be related to essential hypertension, in which inhibited expression of rage is found to be a crucial conflict (Alexander 1939a; 1939b; Saul 1939). Dependent help-seeking tendencies seem to have a close relationship to gastrointestinal activity, as is found in many duodenal ulcer patients (Alexander 1947; Alexander et al. 1934; Van der Heide 1940).

The psychosomatic concept. Although the clinical and experimental study of the mind-body relationship is a comparatively recent development in medicine, the concept is one that dates back to antiquity. The term "psychosomatic" was first used by Johann Christian August Heinroth (1818, part 2, paragraph 313, p. 49), who regarded the body and soul as one, and madness as a disease of the entire being. Christian Friedrich Nasse stated in 1838: "The business of recognizing, preventing, and treating conditions of mental disorder rests upon the fundamental investigation of the simultaneously psychic and somatic activity of man. Here it finds its scientific support, from here on it gains light and learns the road" (in Overholser 1948, p. 231). This statement, echoing an ancient principle of the organism-as-a-whole, is still accepted as an operational concept.

For clarification, one must differentiate the psychosomatic approach, process, disorders, and ill-

nesses. Many reactions, disorders, or illnesses cannot be fully understood unless the investigator can understand the individual's total situation. The *psychosomatic approach* includes both the philosophical orientation to the whole individual and the methodological means of getting data in order to assess reactions to present and past stresses. It attempts to obtain the data—genetic, physical, psychological, and sociocultural—necessary to interpret the forces affecting the life of the individual. The investigator uses varied techniques of data gathering, including interviews, tests, clinical examinations, and laboratory studies.

The *psychosomatic process* refers to that temporal sequential chain of events which occurs in the individual. The end result may be a transitory functional, physiological, and emotional response (*psychosomatic disorder*) that does not result in alteration of the basic organic structure of the individual, or the end result may be a particular syndrome that shows temporary or permanent organic changes (*psychosomatic illness or disease*).

*Psychosomatic medicine* has come into general clinical use with the reintroduction of the word "psychosomatic" by Felix Deutsch, Viktor von Weizsäcker, Helen Flanders Dunbar, Franz Alexander, and others, all of whom stressed the necessity for considering the individual in his totality, and not just as a composite of separate entities and organs (e.g., mind-body).

Deutsch introduced the psychosomatic concept into psychoanalysis by recognizing the influence of repressed emotional conflicts and the unconscious on the functioning of an organ. Deutsch explained that psychosomatic symptoms are the end products of long-existing psychic dynamic processes that always have their origins in past disorders. He demonstrated that the dysfunction of the organic process disappears when unconscious conflicts are made conscious. He defined psychosomatic medicine as the "systematized knowledge of how to study and treat organ processes that are associated and amalgamated with the emotional processes." Deutsch (1964) postulated that the dissolution of psychosomatic symptoms always requires treatment not of the symptoms but of the underlying illness, including the unconscious psychic conflict associated with it.

In the instance of psychosomatic disorders, the notion of a single cause for each disturbance is no longer tenable; in each patient many factors—acting singly or in combination—play significant roles. The psychosomatic approach seeks to identify and understand these variables, their relationships, and their specific contributions toward upsetting the state of equilibrium.



**Homeostasis.** The essential feature of homeostasis is the tendency to maintain a steady state or a state of equilibrium at a given time. Homeostasis is a property of psychological, emotional, and physiological processes, and involves the interplay of various mechanisms. The organism has a constant internal environment, first conceptualized by Claude Bernard, as well as an external environment. Since organisms are functionally indivisible, they cannot be split into the conventional compartments that reflect categories of specialization. Organism and environment form an inseparable pair in dynamic equilibrium. The biological internal environment is the result of natural selection and evolutionary processes. The psychological internal environment, aside from its biological substratum, is the product of the individual's personality development and reflects the sociocultural milieu in which he lives. Although the psychological predisposition to particular patterns of stress reactions may result in part from this psychogenetic developmental process, adaptation may allow various forms of stress and tension to be handled by means of different defense or coping mechanisms and with different pathological results.

W. B. Cannon's concept of homeostasis (1932) clearly emphasized the features of the organization of living systems whereby they tend to maintain themselves as functioning, complete organisms. His pioneering work on the relation of emotions to bodily changes set the theoretical and experimental stage for many of the later clinical and experimental studies in the field of psychosomatic medicine. [See HOMEOSTASIS; and the biography of CANNON.]

**Stress, strain, and stimulus.** Harold Wolff (1953) has considered stress as the internal or resisting force brought into action by external forces or loads. The change in size or shape of an entity as a result of the application of external force is called strain or deformation. Stimuli or external environmental agents are loads; they may be static and sustained, repeated or of brief impact with high intensity. The interaction between external environment and organism is stress. Strain is the alteration or deformation in the organism that ensues. The magnitude of the deformation and the capacity of the organism to withstand the strain determine whether or not there will be re-establishment of homeostasis or a breakdown, with disruption, disorder, disease, and finally death.

Unlike Cannon, Wolff views affect and bodily changes not as being causally connected but as being separate manifestations of response to stimulus, tempered by previous experience. Bernard saw

disease as the outcome of attempts at homeostasis in which adaptive responses to noxious forces, although appropriate in kind, are nonetheless deficient. Wolff suggests that as man is confronted by threats, especially those which involve his values and goals, he initiates responses inappropriate in kind as well as in magnitude. Such reactions, appropriate to one protective purpose, may be inappropriately used for another and can, when inappropriate, be damaging or destructive to the individual.

Threats and symbols of danger may call forth emotional and physiological reactions which in vulnerable individuals differ little from the responses to actual physical assault. Freud first called attention to the signal function of anxiety as an internal reaction to such internal or external dangers. Thus, the stress response resulting from a situation (load) is based in part on the way the affected person perceives the stimulus and the conflict it then sets in motion. Perception and the subsequent interpretation of what is perceived may be dependent upon many factors, including genetically determined responsivity, basic needs and longings, conditioning experiences during formative years, parental and sibling relationships, and identification patterns, as well as other experiences having familial and sociocultural determinants. [See ANXIETY.]

Stress disorders can be expressed psychologically and physiologically. The human organism reacts actively to stress in life situations. The proneness to stress response, the sensitivity and level of threshold, and the specific mode of response are the end products of many factors. Freud described this continuum of factors from the biological to the emotional and sociocultural as a "complementary series." The group of illnesses in which stress seems to be a principal factor has been called stress disorders, in order to avoid using the term "psychosomatic." The form of mental and bodily response is determined by the biological diathesis, the personality pattern, and the character of the situation. When the source of the stress is outside of conscious awareness, or when action is blocked by inner censorship or outward restraints, disharmony or a nonhomeostatic state results, and symptoms of illness appear. It must be noted that stress is the reaction inside the organism, not the force acting on it from the outside. Thus, the inner psychological and physiological dynamics and economics of the individual, as well as the prior and current social and cultural factors, seem to be significant. There are complex intrapsychic phenomena behind every illness, as well as metabolic,

physiological, and even anatomical alterations. These may relate to underlying unconscious fantasies as well as to the emotional significance that different parts of the body and their functions have for the patient. They may be regressions under stress to earlier modes of operation or reflective of earlier fixations. [See STRESS.]

While clinical impressions suggest that there is a specific relation between the psychological stress and the physical disorder that follows, it is not clear how this transformation occurs. The World Health Organization report on psychosomatic disorders (1964, p. 9) summarizes the prevailing ideas on the sequence of involvement as follows:

- (a) constitutional predisposition based on heredity;
- (b) constitutional predisposition laid down as a result of early experience and development (both physiological and psychological experience, and the prenatal period as well as infancy, are included here);
- (c) personality changes of later life that affect organ systems;
- (d) the weakening of an organ, as by an injury or infection;
- (e) the fact that an organ is in action at the moment of strain or emotional upheaval;
- (f) the symbolic meaning of the organ in the personality system of the individual;
- (g) organ-fixation as a result of arrested psychological development.

Empirical, clinical, demographic, and laboratory experimental research involving man and other animals has investigated the effects of psychological stress in symptom-free subjects, as well as in patients suffering from such diseases as duodenal ulcer, ulcerative colitis, bronchial asthma, neurodermatitis, thyrotoxicosis, essential hypertension, rheumatoid arthritis, *diabetes mellitus*, glaucoma, migraine, and coronary heart disease. Clinical studies of human patients have revealed the temporal relation between the onset and the exacerbation of the patient's disease with characteristic emotional upsets. When such patients were treated intensively, e.g., with psychoanalysis, and over a long period of time, much information was obtained, and higher-level abstractions could be formulated for the psychological patterns observed. Some of these empirical clinical studies were published in a volume by Alexander and his associates (1948).

The differentiability of the formulations for seven of these diseases is discussed in *Psychosomatic Specificity* (1968). In one such study, two matched teams of psychoanalysts and internists studied identical interview protocols obtained from patients having one of seven diseases (peptic ulcer, ulcerative colitis, bronchial asthma, neuroderma-

titis, essential hypertension, rheumatoid arthritis, or thyrotoxicosis). These interviews, carefully edited to remove all possible references or cues to the disease of the patient, were evaluated by the two groups from the point of view of formulation and diagnosis of the disease. From the primary clinical information, formulations were prepared which were to explain how psychosocial and emotional conditions predisposed, precipitated, and perpetuated the disease. These comparative diagnoses (by analysts and internists) were statistically evaluated and compared. The results indicate that by and large it was possible to distinguish most of the seven diseases correctly on the basis of psychological abstractions. These abstractions had earlier been derived from the empirical observations of patients seen diagnostically or for therapeutic purposes. These formulations are now being tested by additional studies of families of patients who have psychosomatic diseases, by studying children who have psychosomatic diseases, by detailed investigation of the therapeutic interactions of patients with these illnesses, and by newer laboratory and experimental researches utilizing artificial stresses, such as films, and recording physiological and psychological data following exposure to the conflict situations.

Experimental studies of patients and control groups have been able to reproduce and verify some aspects of the relation between emotional stimuli and the pathophysiological changes of particular diseases. Such studies do not, however, demonstrate pathogenesis in the total etiological sense. What they do show is that particular emotional stresses in sensitive, vulnerable, and predisposed individuals may activate pathophysiological mechanisms which can eventually give rise to anatomical alteration and, thence, to characteristic patterns of disease syndromes. (See Alexander 1950 for discussion of such specific studies.) The precipitating or aggravating situations have specific emotional meaning for the patient because of their connection to his earlier life experiences, his personality development and structure, and his unresolved conflicts that may have been compensatorily handled in various ways until the external situation resulted in too great a stress, with a resulting internal disequilibrium. In this regard, one must recognize that decompensation of adaptive mechanisms may result from too great external demands that cannot be successfully mediated internally, or from a breakdown or weakening of internal control or integrating mechanisms, making the individual more susceptible to disequilibrium effects.



### Major developments

Three major developments occurred early in the twentieth century: the monumental discoveries and theories of Sigmund Freud; the classic and basic physiological and neuropsychological experiments and theories of Walter Cannon; and Ivan Pavlov's research and studies of conditioned reflexes. Contributions were made to the psychosomatic field at two theoretical levels: the first sought to construct general theories and to derive universal principles and concepts, while the second dealt with theories related to studies of particular diseases in which psychosomatic stress is of major significance. Research has continued in both areas. The second level of somewhat limited theory formation offers an opportunity for testing the explanatory aspects of more general and universal theories of psychosomatic functioning. The psychoanalytic and physiological contributions will be discussed here in fuller detail. [See LEARNING, *article on CLASSICAL CONDITIONING*; and the *biography of PAVLOV*.]

**Psychoanalysis.** Psychoanalysis contributed to the understanding of psychosomatic relationships by providing a procedure—free association—which made it possible to study sequences connecting psychological and physiological phenomena. It included the previously undiscovered linkage of unconscious ideas. In addition to improved observation and data gathering, psychoanalytic theory provided a plausible means of reasoning and of understanding previously unintelligible and irrational phenomena, such as dreams and neurotic symptoms. The psychoanalytic method not only permitted the reconstruction of unconscious motivational links and conflicts, but also allowed them to be brought to consciousness and, thus, to be used therapeutically.

Freud's discovery of unconscious mental phenomena and the process of repression threw new light on many bodily expressions of mental phenomena. These mental tensions can be discharged somatically into (1) muscular activities leading to a change of the body in relation to its environment; (2) laughter, crying, screaming, or speech; (3) respiratory, cardiovascular, and other visceral systems. The end effects of these modes of discharge are subjectively experienced as feelings, affects, or emotions. Affects and emotions can be repressed and then are no longer experienced as such, although the processes discharging them into specific somatic systems can still occur. In his study of the unconscious, Freud discovered that strongly charged but repressed fantasies, conflicts, and memories found distorted expression in somatic symptoms

and in impairments of somatic functions. It was discovered that these symptoms were expressions, in symbolic body language, of psychological conflicts. They were called "conversion symptoms": the conflict was "converted" from a purely psychological one into a disorder manifesting itself primarily through somatic symptoms which had no organic pathological correlate. Thus one could see hysterical paralysis, hysterical amnesia, hysterical vomiting, etc. [See HYSTERIA.]

According to Freud, the physiological manifestations of anxiety have a psychological meaning. In his early works he wrote that they consist of a repetition of the physiological manifestations which occurred at the time of the individual's birth. In later writings Freud likened anxiety to a signal used to alert the individual to danger from either internal or external sources. Thus, later threats to survival would evoke physiological responses which occurred in the earliest stress situation, the birth process. [See the *biography of RANK*.]

Hysterical conversion symptoms should not be called psychosomatic disorders, for they do not result in pathological organic processes or lesions, even though they involve the body. They can, however, result in secondary organic disturbances. For example, if a hysterical paralysis of the arms or legs persists, there will be progressive atrophy of the involved limbs because of muscular inactivity.

At its outset psychoanalysis had only a single aim, that of understanding something of the nature of what were then known as the "functional" nervous diseases. The neurologists, with their high respect for physiochemical and anatomical-pathological facts, sought further to establish an intimate and possibly exclusive connection between certain functions and particular parts of the brain. They were puzzled by the psychical factor and did not seem to understand it. In fact, it was believed unscientific to consider these phenomena without neurological explorations. As late as 1885, when Freud was studying at the Salpêtrière, he found that hysterical paralyses were explained by slight functional disturbances of the same parts of the brain which, when severely damaged, led to the corresponding organic paralyses.

In the 1880s, hypnotism provided convincing proof that striking somatic changes could be brought about solely by mental influences which had been set in motion by the patient and were "unconscious" processes. Hypnosis played an important role in Freud's study of hysteria. Charcot's experiments greatly impressed him, especially when Charcot, by suggesting a trauma under hypnosis, was able artificially to induce paralyses.

Charcot's pupil Pierre Janet was able to show, with the help of hypnosis, that the symptoms of hysteria depended on certain unconscious thoughts (*idées fixes*). Janet attributed to hysteria a supposed constitutional incapacity for holding mental processes together—an incapacity which led to a disintegration (dissociation) of mental life. [See HYPNOSIS; and the biographies of CHARCOT and JANET.]

Josef Breuer was the first to study and treat hysteria with hypnosis. His work with a young woman patient made it possible for the first time to obtain a more complete view of a case of hysteria, see the meanings of the symptoms, and note that these symptoms arose in a situation of mental conflict: an impulse toward action had been suppressed because of contrary moral prohibitions, and symptoms appeared in place of the suppressed action. The patient's emotions, involved in a psychic conflict, precipitated the illness, although the traumatic precipitating causes and all the mental impulses relating to them were lost to the patient's memory. Under hypnosis, memories were recovered and, with the subsequent catharsis of affect, the symptoms disappeared.

Breuer and Freud (1893–1895) stated that in hysteria, blocked affect was changed into an unusual somatic innervation (conversion) and resulted in symptoms. Soon after this publication, Breuer and Freud parted, and Freud went on to develop psychoanalysis. He gave up hypnosis for free association and developed his theories of resistance, transference, repression, mental conflict, unconscious mental functioning, and the nature and development of sexual and aggressive drives, including the importance of infantile sexuality.

Freud also suggested that hypochondria and somatic delusions, and their relationship to the pathological return (regression) to an earlier state of psychological organization, are other clinical entities related to psychosomatic syndromes, even though no demonstrable organic pathology is noted in these disorders, as it is in the conversion hysterias. His descriptions and theories of pain and pleasure, and their relationship to feelings and emotions, provide us with an understanding of states of tension. His writings on anxiety and psychic trauma permit us to understand psychic disorganization, its manifestations, and its signal functions. Thus, from the nosological point of view, Freud introduced, described, and explained three clinical categories that are still important: conversion hysteria, anxiety states, and hypochondria and somatic delusions. His concept of the actual neuroses is also of importance, since it states that

the buildup of tension and toxic products without discharge can affect the organism. Freud noted (1910) that not all somatic changes caused by psychological forces were conversions, just as not all physical symptoms were expressions of specific fantasies or conflict compromises. He pointed out that unconscious attitudes may influence organic functions in a physiological way, without the changes having any definite psychic meaning.

Although Freud did not directly address himself to the somatic aspect of affect, his student Karl Abraham wrote about various anatomical zonal correlates of personality (oral and anal stages), and Sándor Ferenczi dealt more directly with the relationship of emotion and mental functioning to organically demonstrable disease. Ferenczi described (1916) pathoneuroses that supervene as a consequence of actual organic illness or injury. Unlike hypochondria, where demonstrable changes in the organ are absent and have never been present at all, pathoneuroses are found in states of actual insult to bodily integrity. In 1922, Hollós and Ferenczi's *Psychoanalysis and the Psychic Disorder of General Paresis* described psychological reactions following central nervous system syphilis. This diagnosis of pathoneurosis is now more generally used to describe the psychic reactions to organic diseases. Ferenczi also described transference manifestations, having psychosomatic implications, that clearly demonstrate Freud's initial idea of fixation and temporary regression to an earlier state of physiological functioning induced by a current situation. [See the biographies of ALEXANDER; FERENCZI.]

Ferenczi also differentiated organ neuroses from conversion hysteria (1926). In the latter there are mainly symbolic and subjective disturbances involving the body, while in the former initially there are mainly objective and demonstrable difficulties affecting organ functioning. Ferenczi described nervous asthma, upper gastric disturbances, intestinal difficulties, cardiac disorders, headaches, fainting, and seasickness as organ neuroses. He noted that purely organic complaints may leave organ-neurotic disturbances behind after they have healed. These may be considered as "secondary gain" phenomena, a situation in which there are psychological advantages of a neurotic variety secondarily connected with the previously present but now absent organic disease.

Among other early psychoanalysts who pioneered in the psychosomatic field, Georg Groddeck (1923) saw purposive activity in all morbid afflictions. Thus, a cold was seen as the wish to avoid smelling, and gynecological disorders as the unconscious



wish for chastity. Groddeck's theory is perhaps the most extreme representative of a panpsychological orientation. He introduced the term *Es* ("id" or "it") as the unconscious formative principle of all normal and abnormal bodily processes. The early writings of Deutsch also indicate the application of Freud's concept of conversion to all dysfunctions of the body. Smith Ely Jelliffe in the United States also attempted to explain all psychosomatic phenomena as direct expressions of highly specific repressed ideas or fantasies (1939). His work, based on careful clinical studies, began as early as 1916, when he first published a paper on psoriasis. Subsequently he and his associates wrote about psychological aspects of tuberculosis, multiple sclerosis, epilepsy, bone disease, postencephalic disorders, endocrinopathies, and eye disorders. Jelliffe was the first American scientist to apply psychoanalytic concepts to organic diseases.

**Physiological contributions.** Early investigators did not distinguish between (a) hysterical conversions based on changes in sensory organs or organs under voluntary nervous control (Freud's view) and (b) changes in internal vegetative organs which are involved with basic visceral functions and are not involved in expressing details of psychological content and in the communication of ideas, thoughts, and feelings. The vocal apparatus, facial expressions, weeping, laughter, blushing, and motor actions discharge and express specific emotional tensions. The internal organs are involved with various bodily processes involving digestion, respiration, circulation, endocrinology, integration, etc. The heart can only beat, the stomach secrete, the bowels contract, the blood vessels constrict or dilate. Although these internal organs respond to emotions, the range of their reactions is limited. Internal organs do not react to specific repressed ideas such as those which underlie hysterical symptoms, but to general emotional states. Neither do these processes, as is true of conversion symptoms, discharge tensions; instead, they are sustained by them.

Alexander made the important distinction between hysterical conversion symptoms and adaptive changes in vegetative functions stimulated by emotional tensions. His ideas followed the concepts of Cannon, who related emotional states, such as fear and rage, to consequent activated physiological functions involving the digestive, endocrine, and cardiovascular systems.

Interest in psychological factors in organic disease became noticeable in Germany after World War I; the work of Leopold Alkan (1930) is especially noteworthy. He indicated that psychogenic disturbances within the autonomic nervous system

may result finally in organic changes, the morphological mechanisms of which form the last link of an intricate causal chain. He postulated that intrapsychic conflicts may be expressed somatically and can result in organic disease which is not reversible.

Karl Fahrenkamp demonstrated the influence of emotions on fluctuations of blood pressure and presented a strong case for the psychogenic origin of essential hypertension (1926). Viktor von Weizsäcker, strongly influenced by Freudian views, emphasized the influence of emotions upon bodily disturbances and of bodily disease upon the psyche (1925; 1925/1926; 1926; 1933). His clinical presentations demonstrated the emotional components and antecedents of bodily disease. Bergmann (1913a; 1913b) and Westphal (1914) suggested the neurotic origin of duodenal ulcers.

**Conditioning studies.** Much research in psychosomatic phenomena has also been carried out within the framework of conditioning theory. Animal studies of psychosomatic disturbances are now on the increase, and there has been considerable study of conditioning of biological functions in humans (e.g., Hofer & Hinkle 1964; Moutsos et al. 1964).

Much interest has been shown in patterns of autonomic responses in normal persons and in patients suffering from psychosomatic disorders, at rest and under stress. Stress can be induced through hypnotic suggestion, interviews, "staged" situations, various test situations—including movies—and by total or partial perceptual isolation. Its effects can be assessed. Physiological variables studied include skin potential, skin temperature, finger pulse volume, cardiac and respiratory rates, blood pressure, and muscle activity. Individual differences in various emotional states, and differences between various psychosomatic disorders, have been reported.

Conditioning experiments demonstrating the effects of anxiety have been carried out on humans and animals (see Dykman et al. 1962; Edwards & Acker 1962; Moss & Edwards 1964; Porter et al. 1958). Predictive studies on the relationship between ovarian activity and psychological processes have been made by Benedek and Rubenstein (1942); on the relationship between emotional states and thyroid function by Dongier and his colleagues (1956); on the etiology of peptic ulcer by Mirsky and his colleagues (1952) and Weiner and his colleagues (1957); and on the appearance of bronchial asthmatic symptoms by Knapp (1963).

Cultural studies indicate that with the extension of Western civilization into rural areas little touched by such acculturation, there is a rising frequency of psychosomatic disturbances (e.g.,

Collomb 1964; Yap 1951). Contrary evidence has appeared about the existence of culture-specific factors determining the choice of symptoms and of disease.

GEORGE H. POLLOCK

[Directly related are the entries HYSTERIA; MENTAL DISORDERS, article on BIOLOGICAL ASPECTS; PAIN; STRESS. Other relevant material may be found in EMOTION; GESTALT THEORY; ILLNESS; PSYCHIATRY; PSYCHOANALYSIS; and in the biographies of ALEXANDER; FREUD.]

# BIBLIOGRAPHY

- ALEXANDER, FRANZ 1939a Emotional Factors in Essential Hypertension. *Psychosomatic Medicine* 1:173-179.
- ALEXANDER, FRANZ 1939b Psychoanalytic Study of a Case of Essential Hypertension. *Psychosomatic Medicine* 1:139-152.
- ALEXANDER, FRANZ 1947 Treatment of a Case of Peptic Ulcer and Personality Disorder. *Psychosomatic Medicine* 9:320-330.
- ALEXANDER, FRANZ 1950 *Psychosomatic Medicine: Its Principles and Applications*. New York: Norton.
- ALEXANDER, FRANZ et al. 1934 The Influence of Psychological Factors Upon Gastro-intestinal Disturbances: A Symposium. *Psychoanalytic Quarterly* 3:501-588.
- ALEXANDER, FRANZ et al. 1948 *Studies in Psychosomatic Medicine: An Approach to the Cause and Treatment of Vegetative Disturbances*. New York: Ronald Press.
- ALKAN, LEOPOLD 1930 *Anatomische Organkrankheiten aus seelischer Ursache*. Stuttgart (Germany): Hippokrates-Verlag.
- ALTSCHULE, MARK D. 1953 *Bodily Physiology in Mental and Emotional Disorders*. New York: Grune.
- APLEY, JOHN; and MACKETH, RONALD 1962 *The Child and His Symptoms: A Psychosomatic Approach*. Philadelphia: Davis.
- ASSOCIATION FOR RESEARCH IN NERVOUS AND MENTAL DISEASE 1939 *The Inter-relationship of Mind and Body*. Research Publications, Vol. 19. Baltimore: Williams & Wilkins.
- ASSOCIATION FOR RESEARCH IN NERVOUS AND MENTAL DISEASE 1950 *Life Stress and Bodily Disease*. Research Publications, Vol. 29. Baltimore: Williams & Wilkins.
- BENEDEK, THERESE; and RUBENSTEIN, BORIS 1942 *The Sexual Cycle in Women: The Relation Between Ovarian Function and Psychodynamic Processes*. *Psychosomatic Medicine Monographs*, Vol. 3, Nos. 1-2. Washington: National Research Council.
- BERGMANN, GUSTAV VON 1913a Das Spasmogene Ulcus pepticum. *Münchener medizinische Wochenschrift* 60:169-174.
- BERGMANN, GUSTAV VON 1913b Ulcus duodeni und vegetatives Nervensystem. *Berliner klinische Wochenschrift* 50:2374-2379.
- BERGMANN, GUSTAV VON 1927 Zum Abbau der "Organneurosen" als Folge interner Diagnostik. *Deutsche medizinische Wochenschrift* 53:2057-2060.
- BREUER, JOSEF; and FREUD, SIGMUND (1893-1895) 1962 *Studies on Hysteria*. Volume 2, pages 3-305 in *The Standard Edition of the Complete Psychological Works of Sigmund Freud*. London: Hogarth; New York: Macmillan. → First published as *Studien über Hysterie*.
- BYKOV, KONSTANTIN M. (1942) 1957 *The Cerebral Cortex and the Internal Organs*. Translated and edited by W. Horsley Gantt. New York: Chemical Publishing. → First published as *Kora golovnogo mozga i vnutrennie organy*.
- CANNON, WALTER B. (1915) 1953 *Bodily Changes in Pain, Hunger, Fear and Rage: An Account of Recent Researches Into the Function of Emotional Excitement*. 2d ed. Boston: Branford.
- CANNON, WALTER B. (1932) 1963 *The Wisdom of the Body*. Rev. & enl. ed. New York: Norton.
- COLLOMB, H. 1964 Psychosomatic Conditions in Africa. Unpublished manuscript, Dakar, Senegal. → An abstract was published in 1964 in *Transcultural Psychiatric Research*, Volume 1, pages 130-134.
- DEUTSCH, FELIX 1949 *Applied Psychoanalysis: Selected Objectives of Psychotherapy*. New York: Grune.
- DEUTSCH, FELIX (editor) 1953 *The Psychosomatic Concept in Psychoanalysis*. New York: International Universities Press.
- DEUTSCH, FELIX 1964 Training in Psychosomatic Medicine. *Advances in Psychosomatic Medicine* 4:35-46.
- DONGIER, M. et al. 1956 Psychophysiological Studies in Thyroid Function. *Psychosomatic Medicine* 18:310-323.
- DUNBAR, HELEN FLANDERS (1935) 1954 *Emotions and Bodily Changes: A Survey of Literature on Psychosomatic Interrelationships, 1910-1953*. 4th ed. New York: Columbia Univ. Press.
- DUNBAR, HELEN FLANDERS 1943 *Psychosomatic Diagnosis*. New York: Hoeber.
- DUNBAR, HELEN FLANDERS 1947 *Mind and Body: Psychosomatic Medicine*. New York: Random House.
- DUNBAR, HELEN FLANDERS 1948 *Synopsis of Psychosomatic Diagnosis and Treatment*. St. Louis, Mo.: Mosby.
- DUNBAR, HELEN FLANDERS 1959 *Psychiatry in the Medical Specialties*. New York: McGraw-Hill.
- DYKMAN, ROSCOE A. et al. 1962 Inhibition of Urine Flow as a Component of the Conditional Defense Reaction. *Psychosomatic Medicine* 24:177-186.
- EDWARDS, ALLAN E.; and ACKER, LOREN E. 1962 A Demonstration of the Long-term Retention of a Conditioned GSR. *Psychosomatic Medicine* 24:459-463.
- FAHRENKAMP, KARL 1926 *Die psycho-physischen Wechselwirkungen bei den Hypertonieerkrankungen: Eine klinische Studie über die praktische Bedeutung der Blutdruckkurve*. Stuttgart (Germany): Hippokrates-Verlag.
- FERENCZI, SÁNDOR 1916/1917 Von Krankheits- oder Pathoneurosen. *Internationale Zeitschrift für Psychoanalyse* 4:219-228.
- FERENCZI, SÁNDOR 1923 Ptyalismus bei Oralerotik. *Internationale Zeitschrift für Psychoanalyse* 9:67 only.
- FERENCZI, SÁNDOR (1926) 1964 *Organ Neurosen und ihre Behandlung*. Volume 3, pages 294-301 in Sándor Ferenczi, *Bausteine zur Psychoanalyse*. 2d ed. Stuttgart (Germany): Huber.
- FRENCH, THOMAS M. et al. 1941 *Psychogenic Factors in Bronchial Asthma*. *Psychosomatic Medicine Monographs*, Vol. 1, No. 4; Vol. 2, Nos. 1-2. Washington: National Research Council.
- FREUD, SIGMUND (1910) 1957 *The Psycho-analytic View of Psychogenic Disturbance of Vision*. Volume 11, pages 209-218 in *The Standard Edition of the Complete Psychological Works of Sigmund Freud*. London: Hogarth; New York: Macmillan. → First published in German.
- GREGG, ALAN 1937 *The Future of Medicine*. *Federation Bulletin* 23:5-12.



- GRINKER, ROY R. 1953 *Psychosomatic Research*. New York: Norton.
- GRINKER, ROY R.; and ROBBINS, FRED P. 1954 *Psychosomatic Case Book*. New York: Blakiston.
- GRODECK, GEORG W. (1923) 1950 *The Book of the It*. New York: Funk & Wagnalls. → First published as *Das Buch vom Es*.
- HEINROTH, JOHANN C. A. 1818 *Lehrbuch der Störungen des Seelenlebens oder der Seelenstörungen und ihrer Behandlung*. Leipzig: Vogel.
- HOFER, MYRON A.; and HINKLE, LAWRENCE E. JR. 1964 Conditioned Diuresis in Man: Effects of Altered Environment, Subjective State and Conditioning Experience. *Psychosomatic Medicine* 26:108-124.
- HOLLÖS, STEFAN; and FERENCZI, SÁNDOR (1922) 1925 *Psychoanalysis and the Psychic Disorder of General Paresis*. Nervous and Mental Disease Monograph No. 42. New York: Nervous and Mental Disease Monograph Publishing. → First published in German.
- JELLIFFE, SMITH E. 1939 *Sketches in Psychosomatic Medicine*. Nervous and Mental Disease Monograph No. 65. New York: Nervous and Mental Disease Monograph Publishing.
- KNAPP, P. H. 1963 Short-term Psychoanalytic and Psychosomatic Predictions. *Journal of the American Psychoanalytic Association* 11:245-280.
- LIEF, HAROLD I. et al. (editors) 1963 *The Psychological Basis of Medical Practice*. New York: Harper.
- MARGETTS, EDWARD L. 1950 The Early History of the Word "Psychosomatic." *Journal of the Canadian Medical Association* 63:402-404.
- MASSACHUSETTS GENERAL HOSPITAL, BOSTON, PSYCHIATRIC SERVICE 1952 *Case Histories in Psychosomatic Medicine*. Edited by H. H. W. Miles, Stanley Cobb, and Harley C. Shands. New York: Norton.
- MENTAL HEALTH RESEARCH FUND, CONFERENCE, SECOND, OXFORD, 1958 1960 *Stress and Psychiatric Disorder: The Proceedings*. Edited by J. M. Tanner. Oxford: Blackwell.
- MIRSKY, I. A. et al. 1952 Blood Plasma Pepsinogen: The Activity of the Plasma From "Normal" Subjects, Patients With Duodenal Ulcer and Patients With Pernicious Anemia. *Journal of Laboratory and Clinical Medicine* 40:188-199.
- MOSS, THELMA; and EDWARDS, ALLAN E. 1964 Conflict vs. Conditioning: Effects Upon Peripheral Vascular Activity. *Psychosomatic Medicine* 26:267-273.
- MOUTSOS, SPERO E. et al. 1964 A Study of Conditioning and Reactivity as Determinants of Pressor Responses to Noxious Stimuli. *Psychosomatic Medicine* 26:274-286.
- NODINE, JOHN H.; and MOYER, JOHN H. (editors) 1962 *Psychosomatic Medicine: The First Hahnemann Symposium*. Philadelphia: Lea & Febiger.
- OVERHOLSER, WINFRED 1948 *Psychiatry and General Medicine*. *Mental Hygiene* 32:226-234.
- PORTER, R. W. et al. 1958 Some Experimental Observations on Gastro-Intestinal Lesions in Behaviorally Conditioned Monkeys. *Psychosomatic Medicine* 20:379-394.
- Psychosomatic Specificity*. Volume 1: Experimental Study and Results. Edited by Franz Alexander, Thomas M. French, and George H. Pollock. 1968 Univ. of Chicago Press.
- RUESCH, JURGEN (1909) 1946 *Chronic Disease and Psychological Invalidism: A Psychosomatic Study*. New York: American Society for Research in Psychosomatic Problems.
- SAUL, LEON J. 1939 Hostility in Cases of Essential Hypertension. *Psychosomatic Medicine* 1:153-161.
- SCHNEER, HENRY I. (editor) 1963 *The Asthmatic Child: Psychosomatic Approach to Problems and Treatment*. New York: Harper.
- SELYE, HANS 1956 *The Stress of Life*. New York: McGraw-Hill.
- SOCIETY FOR PSYCHOSOMATIC RESEARCH 1959 *The Nature of Stress Disorder*. Springfield, Ill.: Thomas.
- SOCIETY FOR PSYCHOSOMATIC RESEARCH 1965a *Psychosomatic Disorders in Adolescents and Young Adults*. Edited by John Hambling and Philip Hopkins. Oxford and New York: Pergamon.
- SOCIETY FOR PSYCHOSOMATIC RESEARCH 1965b *The Role of Psychosomatic Disorder in Adult Life*. Edited by John O. Wisdom and H. H. Wolf. Oxford and New York: Pergamon.
- VAN DER HEIDE, CAREL 1940 A Study of Mechanisms in Two Cases of Peptic Ulcer. *Psychosomatic Medicine* 2:398-410.
- WEINER, HERBERT et al. 1957 Etiology of Duodenal Ulcer. *Psychosomatic Medicine* 19:1-10.
- WEISS, EDWARD; and ENGLISH, O. SPURGEON (1943) 1957 *Psychosomatic Medicine: A Clinical Study of Psychophysiological Reactions*. 3d ed. Philadelphia: Saunders. → The 1943 edition was subtitled *The Clinical Application of Psychopathology to General Medical Problems*.
- WEIZSÄCKER, VIKTOR VON 1925 Randbemerkungen über Aufgabe und Begriff der Nervenheilkunde. *Deutsche Zeitschrift für Nervenheilkunde* 87:1-22.
- WEIZSÄCKER, VIKTOR VON 1925/1926 Über neurotischen Aufbau bei inneren Krankheiten. *Deutsche Zeitschrift für Nervenheilkunde* 88:264-271.
- WEIZSÄCKER, VIKTOR VON 1926 Der neurotische Aufbau bei den Magen- und Darmerkrankungen. *Deutsche medizinische Wochenschrift* 52:2103, 2105, 2150-2152.
- WEIZSÄCKER, VIKTOR VON 1933 Körpergeschehen und Neurose: Analytische Studie über somatische Symptombildungen. *Internationale Zeitschrift für Psychoanalyse* 19:16-116.
- WESTPHAL, KURT 1914 Untersuchungen zur Frage der nervösen Entstehung peptischer Ulcera. *Deutsches Archiv für klinische Medizin* 114:327-395.
- WESTPHAL, KURT; and BAR, R. 1926 Über die Entstehung des Schlaganfalles. *Deutsches Archiv für klinische Medizin* 151:1-109.
- WESTPHAL, KURT; and WELT, M. H. 1930 Tuberkulose und Schizophrenie. *Klinische Wochenschrift* 9:1025-1026.
- WHITE, BENJAMIN et al. 1939 *Mucous Colitis: A Psychological Medical Study of Sixty Cases*. Washington: National Research Council.
- WITTKOWER, ERIC D.; and CLEGHORN, R. A. (editors) 1954 *Recent Developments in Psychosomatic Medicine*. Philadelphia: Lippincott.
- WOLFF, HAROLD 1953 *Stress and Disease*. Springfield, Ill.: Thomas.
- WORLD HEALTH ORGANIZATION, EXPERT COMMITTEE ON MENTAL HEALTH 1964 *Psychosomatic Disorders: Thirteenth Report of the WHO Expert Committee on Mental Health*. World Health Organization, Technical Report Series No. 275.
- YAP, P. M. 1951 Mental Diseases Peculiar to Certain Cultures: A Survey of Comparative Psychiatry. *Journal of Mental Science* 97:313-327.

## PSYCHOTHERAPY

See MENTAL DISORDERS, TREATMENT OF; PSYCHOANALYSIS.

## PUBLIC ADMINISTRATION

The process of public administration consists of the actions involved in effecting the intent or desire of a government. It is thus the continuously active, "business" part of government, concerned with carrying out the law, as made by legislative bodies (or other authoritative agents) and interpreted by courts, through the processes of organization and management.

The field of study—putatively a science or discipline—of Public Administration focuses upon public administration as a process. (In this article, *public administration* is used to refer to the relevant governmental phenomena; *Public Administration*, to refer to the study of these phenomena.) Historically and conventionally, Public Administration has been primarily concerned with problems of how to apply or effect law faithfully, honestly, economically, and efficiently. More recently, Public Administration has become concerned with the processes by which public administration participates in creating and interpreting law—with how such creating and interpreting can be done "correctly," "wisely," or in the "public interest."

As a process, public administration is as old as government. That is, as soon as there is sufficient institutional evolution and differentiation to enable one to speak of the government of a society, there are actions by which law (as an expression of government's authoritative allocation of values) is made and actions by which an attempt, more or less successful, is made to carry the law into effect. In simple societies the objectives of public administration are simple, limited to such matters as preserving order; and the institutions or organs by which administration is carried on are simple, comparatively small, and often not completely differentiated from institutions or organs with other purposes. As societies increase in size and complexity, as governments grow larger and take on more functions, and as their institutions become more differentiated and specialized, administrative processes become more specialized and the institutions that carry on administration activities—known by such names as councils, commissions, departments, bureaus, and agencies—become large, complex, and highly differentiated.

As public administration becomes more specialized and complex, increasing attention is likely to

be given to the training of persons who are to perform administrative functions, training which may be given either before or after entry into service. As the processes of administration become more complex, considerable attention also may be given to means of improvement of some part or process. Systematic education or training for performance in public administration, however, is for the most part a development of the modern era. And systematic, continuous study of ways to improve public administration and make it more efficient is an even more recent development, associated with various modern developments such as the rise of the nation-state as the dominant governmental form and the rise of science as an acceptable or characteristic way of thought. In the United States during the past generation the study of public administration has been especially intense, and this intensity of effort has resulted in a new level of self-consciousness. This is reflected in the idea that the study of public administration is sufficiently important and sufficiently autonomous to become a science or discipline in itself. The conception of a more or less autonomous science or discipline called Public Administration is primarily, if not indeed uniquely, an American idea.

**Terminological difficulties.** A note on terminology is in order. It is difficult to be either brief or accurate—and more difficult to be both—in speaking of Public Administration in the United States. The key terms, such as *administration*, *execution*, and *management*, have no precise and agreed meanings but rather overlap and conflict, and the differences in usage relate not simply to carelessness and accident but to matters of disciplinary doctrine and methodological and professional dispute. Centrally relevant is the fact that the U.S. constitution has established a threefold separation of institutions and organs, legislative, executive, and judicial, yet relates the three in a complicated fashion in which each of the three branches has functions that in a logic of strict separation would belong to the other two; and this general scheme of construction is repeated in the constitutions of the constituent states. Moreover, the word *administration* does not occur in the federal constitution. Thus the general question of how *administration* (or *management*) relates to *executive* as used in the constitution—particularly, what institutions and persons are to direct and control administration—is open to dispute. Students in Public Administration have traditionally had and expressed an antilegal bias. The preface of the first Public Administration textbook explicitly states the thesis that "the study of administration should start from



the base of management rather than the foundation of law" (White 1926). It is vitally necessary to understand that the rise of Public Administration represented an assertion that the traditional view of public administration as simply the application of legal rules was quite inadequate.

### The field in the United States

In the 1950s and 1960s Public Administration changed very greatly. But the nature and significance of these changes can only be understood in terms of past doctrines and interests. For this it is necessary to sketch out the historical and institutional context for the rise of the discipline.

The first two general textbooks of Public Administration in the United States, written by L. D. White and W. F. Willoughby, were published in 1926 and 1927 respectively. In a sense these signified the birth of Public Administration as a discipline. Before these general textbooks appeared, however, there had been several decades of preparation.

**Origins.** Some of the framers of the U.S. constitution and some early U.S. political leaders—for example, Alexander Hamilton and Thomas Jefferson—gave attention to problems of public administration and wrote on them in ways that foreshadowed later developments. Indeed, there is no sharp point in history where the story of Public Administration begins. However, an essay by the then young Woodrow Wilson (1887) is often taken as the symbolic beginning. Certainly it was a remarkable essay in its perceptiveness, persuasiveness, and influence [see the *biography* of WILSON].

Wilson's basic postulate was that "it is getting to be harder to run a constitution than to frame one" ([1887] 1953, p. 67). Up to the nineteenth century, he noted, the predominant concerns of the study of governmental affairs were political philosophy, constitutional arrangements, and lawmaking. However, increasing complexities in economic and social life and a concomitant increase in governmental size and activity were forcing a change of emphasis. European countries, he observed, had begun taking very seriously the training of civil servants and the scientific study of administration. The United States should do likewise. In fact, it should study European methods, to learn from them—taking care in the borrowing of these efficient means not to borrow their monarchical or autocratic ends. The United States should rather seek to perfect its republican-democratic constitutional system through the new science: "There should be a science of administration which shall seek to straighten the paths of government, to

make its business less unbusinesslike, to strengthen and purify its organization, and to crown its duties with dutifulness" (*ibid.*).

The rise of Public Administration as a discipline, and the reforms and changes in public administration that stimulated it and that it in turn stimulated and guided, must be understood in relation to developments in national life. Public Administration, that is, represented a response to threats to old values, an adjustment to new conditions of life. As a body of thought and techniques it has been an attempt to preserve the essential parts of the republican-democratic heritage, conceived and developed under comparatively simple rural conditions, under the extremely complex conditions posed by a large industrial nation, itself situated in an international setting of increasing complexity.

Also important to the rise of Public Administration was the rise to self-consciousness of the discipline of political science, evidenced by the creation of the American Political Science Association in 1903. It is important that almost without exception those who might be called the founding fathers of Public Administration were trained as political scientists (rather than, say, jurists or economists) and tended to view Public Administration as a part or subdiscipline of political science. American higher education generally—the modern university with its professional schools—was developing swiftly in these decades. Specialization and expertise were replacing amateurism in many realms, and the development of political science was but an aspect of a movement; the other major social sciences also were achieving differentiation and separate status in the same period.

The 1920s and 1930s. Public Administration acquired certain distinctive characteristics in the 1920s and 1930s.

*Public Administration as a synthesis.* As it emerged in the mid-1920s Public Administration was a confluence and mixture of three main currents. One of these was the movement for governmental reform, a movement that had been swelling for several decades, which sought to purify government and adjust its institutions to the conditions of modern, industrial, urban society. The second current was scientific management, or what often came to be called the "management movement": the attempt, centered in private industrial enterprise but extending out from it, to apply the techniques of science to solving problems of organizing and administering with economy and efficiency. The techniques and spirit of scientific management began also to pervade governmental administration and to influence the thinking of those interested

in governmental reform and improvement. Its use of the increasingly honorific term *science*, its promise to reduce cost and increase efficiency, and its aim of replacing ignorance and conflict with knowledge and harmony were extremely appealing to the founding fathers of Public Administration. Scientific management also influenced Public Administration through the concurrently developing discipline of business administration. The third current was the new discipline of political science, which inherited the political philosophies and governmental knowledge of the centuries but was also trying to apply *science* in a new and more rigorous way to politics and government. Other influences—foreign and domestic—are readily recognizable, but these three constitute the essential ingredients.

*The core beliefs.* The new Public Administration of the 1920s had certain core, orienting beliefs. Centrally, its leaders thought of their enterprise as an attempt to achieve the republican-democratic ends of freedom and equality by making government simultaneously strong and efficient, responsible and responsive. A program of political reform as well as a mere science of administration was involved. Against the nineteenth-century idea of dividing powers and sharing functions widely among the citizenry was posed a new formula, necessitated by the need for expertise and by modern conditions generally: concentrate power for effectiveness and then watch it closely for responsibility. Indeed, the concentration is as necessary for achieving responsibility as for effectiveness; for only if the citizen's task is simplified and he knows whom to hold responsible, through his vote and by other means, can he make his influence felt.

In the envisaged reconciliation of "true democracy" and "true efficiency" one postulate is very important, namely, that government is analytically divisible into two different functions, or types of action: first, to decide; and second, to carry out the decision. These two types of action are identified with and by the words *politics* and *administration*, respectively. The role of public opinion, the activities of political parties, the functions of legislative bodies—these are identified with *politics*. Here the clash of opinion and the conflict of values are in order, and science can have only a limited role, efficiency a limited meaning. But once a decision is authoritatively made and a law enacted, then other values and techniques are primarily appropriate; this is the realm of *administration*. For this, economy and efficiency are the central criteria, and science is the proper method for developing the criteria.

This division of government between politics and

administration had many sources—including, again, Wilson's essay—but one especially influential book must be mentioned: Frank Goodnow's *Politics and Administration* (1900). In the idiom of that day, Goodnow spoke of the "will of the state" and identified politics with the *expression* and administration with the *execution* of this will. The problem of government, as he viewed it, is to achieve harmony between the expression of the will and its execution, the alternative being conflict or paralysis. But while politics should have a certain control over administration, it should not extend to certain levels and aspects thereof, which embrace "fields of semi-scientific, quasi-judicial and quasi-business or commercial activity—work which has little if any influence on the expression of the true state will" (p. 85).

Goodnow made recommendations for various reforms to achieve this necessary and desirable, but limited, control of administration by politics. Goodnow himself did not identify politics simply with the legislative organs or the lawmaking process, nor administration simply with the executive branch or the process of law enforcement. However, as time passed, the student of public administration tended to identify his subject—and indeed himself—with executive institutions and processes and to presume that his subject had qualities of "hardness" lacking in politics and policy making, which made it possible to use science as the central method in study and to take efficiency as the central criterion of success in operation. [See the biography of GOODNOW.]

*The period of orthodoxy.* The expression *orthodoxy* is now often used with reference to Public Administration in the 1920s and 1930s to indicate a quality of general agreement and self-assurance. To be sure, there were differences of seemingly great import among the scholars of the day, but in retrospect certain general beliefs and interests predominated.

The early textbooks afford a summary view of the interests of the newly founded discipline. The most widely used textbook in the period was Leonard D. White's *Introduction to the Study of Public Administration* (1926). In the first edition, Chapter 1 is entitled "Administration and the Modern State." Here are set forth some of what have been designated above as core beliefs: the importance of administration in an increasingly complex and interdependent society, the necessity for efficiency, the possibility and desirability of approaching administration scientifically, and so forth.

The second textbook, W. F. Willoughby's *Prin-*



*ciples of Public Administration* (1927), differs significantly from White's in regard to the constitutional authority to control public administration: White regards the president as the chief administrator by constitutional right, whereas Willoughby regards Congress as holding the administrative power by constitutional right, delegating it to the president or other officers at its discretion. In terms of subject matter the chief difference lies in the fact that Willoughby devotes a large part of his book to financial and budgetary aspects of administration and has a section as well dealing with "matériel"—purchasing, storage, and so forth. Textbooks appearing in the 1930s—a revision of White, and others—included treatment of governmental financial-budgetary matters and at least touched lightly on "matériel."

The word *principles* in the title of Willoughby's book indicates an important doctrinal aspect of the older Public Administration that has not yet been mentioned: It was customary to speak of "principles of (public) administration." A principle was what one arrived at by proper and diligent study; it summed up what had been learned. At the same time, it was a guide to good and efficient administration, a light thrown by research and logic on present and future problems.

The core beliefs may now be summarized. Government is divisible into two functions or processes, decision and execution. Making decisions is the realm of politics and policy making. It is the area in which the processes of democracy are relevant—expression of opinion, voting, organization of political parties, and so forth. Executing decisions, however, which is the realm of administration, presents other problems and needs other criteria. To the processes of administration the methods of science, proved so powerful elsewhere, are relevant. The criteria here are economy and efficiency; and economy can on close analysis be viewed as an aspect of efficiency. Through scientific research of the phenomena of administration we can derive principles of administration, which simultaneously summarize what we have learned and provide formulas for the efficient conduct of administration. By this process of analysis and division we can reconcile the values of democracy with the necessities of efficiency and science in the modern world.

**Criticism and transition—the 1940s.** Roughly speaking, the period of so-called orthodoxy in Public Administration coincided with the years between the two world wars. In the 1940s the discipline was subjected to searching criticism of its core beliefs, and heterodoxy came to replace orthodoxy. The

criticism and new orienting ideas were clearly foreshadowed in the 1930s. In 1937 a work appeared that is commonly regarded as the epitome of orthodoxy, *Papers on the Science of Administration*, edited by Luther H. Gulick and Lyndall Urwick. But this work not only represented orthodoxy in its most cogent and influential form; it presented papers dealing with the psychological dimensions of administration, which were to become so important in the postwar years. Moreover, in 1936 there had appeared *The Frontiers of Public Administration*, a series of essays by John M. Gaus, Leonard D. White, and Marshall E. Dimock, in which these prominent figures of the orthodox period introduced points of view later to become important.

Other works of the prewar period criticizing old ideas or introducing new could be cited, but World War II diverted normal streams of scholarly activity, and it was not until the second half of the decade of the 1940s that the most influential critical works appeared. At the same time, however, it should be recognized that the war itself was important in stimulating dissatisfaction with old perspectives and encouraging new ones. Most of the university students and teachers of Public Administration either were engaged in some form of emergency administrative activity or were in some military organization, and they concluded that the textbooks of the day did not adequately describe organization and administration as they had experienced it.

The critical ideas of the younger students were indicated by Robert A. Dahl in 1947, in an essay entitled "The Science of Public Administration: Three Problems." The first problem arises from the "frequent impossibility of excluding normative considerations from the problems of public administration" (p. 1), as had been the intent or tendency with the politics-administration dichotomy and the accompanying focus upon scientific means to achieve efficiency. Dahl argued that we must "recognize that the study of public administration must be founded on some clarification of ends" (p. 3). The second problem arises from the "inescapable fact that a science of public administration must be a study of certain aspects of human behavior" (p. 4). He criticized the prevalent tendency to treat organization in formal, technical terms and to regard the human beings that constitute organizations more or less as "material." The study of administration must, he argued, embrace the whole psychological man and must not presume that man is a simple machine responding only and fully to goals of self-interest narrowly conceived. The third

problem concerns the conception of principles of administration. The study of Public Administration in the United States, he argued, has been too narrow, too parochial. We have hoped, he said, indeed presumed, that we were enunciating universal principles, but our study has, after all, been limited to a few examples in a few national and historical settings, and we presume too much. "The study of public administration inevitably must become a much more broadly based discipline, resting not on a narrowly defined knowledge of techniques and processes, but rather extending to the varying historical, sociological, economic and other conditioning factors . . ." (p. 11).

Herbert A. Simon's *Administrative Behavior: A Study of Decision-making Processes in Administrative Organization* (1947) was probably the most important work of the 1940s. It contained a searching critique of the older Public Administration, particularly of its use of "principles." These so-called principles, he observed, are similar to maxims of folk wisdom and, in fact, given the loose and unscientific way in which they have been derived and stated, cannot be regarded as more than proverbs.

*Administrative Behavior* represents the direct and vigorous impact on Public Administration of the perspectives and methodology associated with behavioralism (a movement in the social sciences aimed at a higher level of achievement by more careful study of actual behavior, using techniques whose value has been demonstrated in the physical sciences).

According to Simon, the founders of the older Public Administration failed to appreciate many of the rigorous requirements of true scientific method, but their *fundamental* deficiency lay in their lack of understanding of the distinctions they had drawn. They failed to appreciate that their rough separation of politics from administration did not preclude a valuational component in many things they presumed they were treating scientifically. In fact, their principles typically represented a conflation and a confusion of the two elements of fact and value.

The philosophical-methodological concerns of the book and the more purely substantive materials (on such subjects as communication and authority) are joined centrally through the concept of *decision making*: "If any 'theory' is involved, it is that decision-making is the heart of administration, and that the vocabulary of administrative theory must be derived from the logic and psychology of human choice" (Simon [1947] 1961, p. xiv). As for the psychology of human choice, Simon selected

what seemed to him relevant theories from psychology and various social sciences.

*Administrative Behavior* was paradoxically a radical and a conservative work with respect to Public Administration. It was radical in its rejection of the politics-administration dichotomy and the simultaneous injection of the perspective of logical positivism in approaching questions of policy making and the relation of means and ends, in its proposal to adopt what was becoming a very fashionable term and set of concepts in the social sciences, *decision making*, into Public Administration; and in its insistence that standards of scientific rigor in Public Administration be sharply raised. At the same time, *Administrative Behavior* was faithful to some essential beliefs of the older Public Administration. At a time when its claim to be a science was under attack as pretentious, Simon argued forcefully that administrative phenomena are indeed the proper subjects of scientific study—if properly conceived and executed. At a time when the concept of efficiency was under criticism as too narrow and unimaginative a criterion, he carefully defined and refined the concept, made a distinction between pure and practical sciences, and argued that efficiency is a proper criterion as applied to the factual aspects of a practical science of administration. Even the sharp distinction between fact and value, while a much more subtle matter than the distinction between politics and administration, resembles the latter in the formal sense that it is a sweeping twofold division of the universe of administrative phenomena.

Main currents of recent years. Since the critical analyses of the 1940s, Public Administration as a discipline has lacked the self-confidence and coherence of the interwar period. Various approaches or emphases have competed, but none has succeeded in winning the general acceptance of scholars identified with the discipline. No new synthesis has been achieved; no new orthodoxy has replaced the old. In general, Public Administration has grown tremendously in the sense of accepting data, concepts, and perspectives from many sources, chiefly the various social sciences; but it has discarded little, and no organizing framework into which everything will fit has been achieved—or, if achieved, has not been recognized and accepted as such. There are several major currents and emphases, but these often overlap and mingle and, while separated for purposes of discussion, are not necessarily separated in particular books, courses of study, and so forth.

*Continuation of the traditional.* First, it should be noted that although the 1940s undoubtedly mark



a period of criticism and even rejection of the discipline as it emerged in the interwar period, still there has been much continuity with the older Public Administration. Scholars and teachers have naturally varied in their rejection of the traditional and their acceptance of the new. New data and concepts were added to the old; traditional points of view were presented—but qualified and criticized. Most teachers, for example, would present the organization theory of the 1930s because of its wide acceptance and practical importance, whatever criticism they might make or whatever other concepts of organization they might introduce in addition. No teacher would speak confidently of principles of administration. Still, faced with teaching students interested in careers in public administration, he would likely find himself falling back upon some of the “managerial,” efficiency-oriented attitudes and materials of his predecessors.

*Politics and policy making.* The belief that the practice of administration is a technical problem focused upon efficiency in operation was characteristic of the older Public Administration. The leading students and writers of the postwar period have adopted sharply different attitudes—whatever their other differences—on this matter. It has been generally agreed that while the phenomena of politics and the amount of policy making may decrease as one moves from the top of an administrative agency to its bottom, or into some of the technical processes or functions, still they are generally present in significant degree; and at the level of chief executive or top management, where so much interest is focused, they are important matters indeed.

The results of this recognition have manifested themselves in a variety of approaches. For example, Simon's decision-making schema attempts to include the valuational as well as the factual. Some writers, for example Paul H. Appleby (1952), have written searchingly on the interaction of politics and administration in a democracy. Some, for example Norton E. Long (1962), have concentrated more sharply on politics-in-administration, on the power factor in administration. Some, for example Emmette Redford (1958), have reflected on how the ethical or public-policy component is brought to bear on the technical component.

*Development of a case method.* One important development of the postwar period, closely related to the abandonment of the politics-administration dichotomy, has been the development of a case method of study and teaching. There had, indeed, been case programs earlier, but in the mid-1940s at

Harvard University there were new beginnings and a development of new objectives and new techniques. An interuniversity program followed the Harvard initiative, and since 1951 the enterprise, further expanded, has operated under the title Inter-University Case Program.

Essentially, those who developed the new case style and format were dissatisfied with older Public Administration as doctrinaire and limited, failing to present and deal adequately with real public administration. As developed, the case is a narrative, an account of a particular, real administrative episode, as written after the events from information gathered from all possible sources. The perspective of the writer might be characterized as “interested but impartial observer,” but he often tries to re-create the perspective of an important participant (or participants) in the episode—a man in a situation having to make a decision. There is an effort to present the “entire situation,” i.e., everything that is relevant to the decision; and the emphasis is more on personal interaction, politics, and policy making than on technical factors.

The production of such cases has been and continues to be a major enterprise of the group of scholars who produce much of the literature in the field. Since in an important sense the focus in case writing is decision making, this might seem to join them to, or at least bring them into relationship with, Simon's proposal to center the study of administration on decision making. In fact, however, serious philosophical-methodological differences have separated the two despite the formal use of the same term.

*“Human relations,” psychology, sociology.* In the late 1920s a series of famous experiments focusing upon work groups was carried out at the Hawthorne plant of the Western Electric Company (see Roethlisberger & Dickson 1939). The impact of these experiments was very great on both the study of business administration and that of public administration, though in the case of the latter the full effect was not felt until the postwar period.

The Hawthorne experiments demonstrated not only the limitations of the scientific-management approach to increased efficiency by concentration upon monetary rewards and the physical aspects of the work situation but also the importance of psychological and, more broadly, social factors. The Hawthorne experiments signaled a vogue: the “human relations” approach to management. More recently, this term has lost its appeal, as associated with certain excesses and naivete. However, the human relations movement set in train a vast

amount of serious scientific research on the psychosociological aspects of work and administration [see INDUSTRIAL RELATIONS].

The questions to which psychological research is directed are, for example: What is morale, what factors affect it, and how is it related to work effectiveness? How do factors in a worker's perceptions and attitudes relate to the stimuli in the work situation? What factors affect the formation of face-to-face work groups, and what is their significance in administration? How does informal group life generally relate to the formal organization structure and to the official organizational goals? What are the factors involved in the phenomenon of leadership, and how do they relate to the important matter of authority? What makes for conflict, what for cooperation, between groups or organizations? [See LEADERSHIP.]

Research in administration by sociologists has become increasingly important during the postwar period. Only a small part of this research has taken public organizations as its focus, but this part has a high scientific relevance; and all of it has a high potential relevance. Sociology has brought many and varied concepts to administration, but most have clustered about the term *bureaucracy*, an ideal type set forth by Max Weber. The research findings of sociology have less obvious relevance to immediate problems of efficiency than do those of psychology, but the range of interests and techniques embraced may hold greater promise for coping with organizational problems as they relate to the whole of society. [See BUREAUCRACY.] In general, however, Public Administration stands to benefit from the research interests and techniques of both psychology and sociology.

*Theory of organization.* In recent years there has been great scholarly and scientific interest in organization, and a resultant outpouring of essays, books, and research reports designated by the term *theory of organization*. It is the view of some that organization itself, as a widespread social phenomenon, warrants the full concentration of many students. There is a widely held belief that there are universals in organizational behavior, that organization can be studied simply as organization, with a resulting body of valid theory of general applicability. One group of students, identified with general systems theory, conceives of human organization as but a type or representative of a still more general phenomenon: systems (which may be biological or physical as well as social) [see SYSTEMS ANALYSIS].

Some students of public administration are con-

tributing to theory of organization; others, not identified with Public Administration, focus their research effort on public organizations. To the extent that there may be universals of organization, they are by definition relevant to public organizations. In any event, theory of organization is one of the very active areas of contemporary social science related to and contributing to Public Administration. [See ORGANIZATIONS.]

*Comparative Public Administration.* Within Public Administration proper, perhaps the area of greatest current scholarly activity—and some would say of greatest promise—is the comparative study of public administration. This interest grows out of the fact that, beginning with World War II, continuing into the postwar military occupations, and accelerating with the many technical assistance programs of the United States, the United Nations, and private foundations, American students and teachers of Public Administration by the hundreds have found themselves engaged in professional work in foreign lands. This exposure to foreign, often non-Western, governmental systems and cultures has stimulated a sense of "comparativeness" in general, and in particular raised questions either about the appropriateness or the sheer possibility of transferring familiar administrative devices or applying what had been presumed to be good or scientific principles of administration.

The comparative study of administrative systems parallels the comparative study of political systems—comparative politics. Both movements are characterized by the comparative youth of their participants, by a general commitment to the outlooks identified with behavioralism, by an effort to be interdisciplinary in interests and techniques, and by an effort to arrive at concepts, formulas, and theories that are truly universal, bridging and embracing all cultures. [See POLITICS, COMPARATIVE.]

*New technologies and techniques.* Recent years have seen the rapid development of various physical devices of significance for administration and a parallel development of various systems of logical thought applicable to the practice of, and to research on, administration. The former include a range of new machines, mostly electronic, that greatly extend the range of—or replace—the hand and, particularly, the brain. These devices, of which the electronic computer in its various forms is the best known, perform certain work, particularly the storage, retrieval, and manipulation of data and the making of complex calculations, with such speed and accuracy that they open up a new range of human experience and potentialities. The sys-



tems of logical thought that have been newly developed for or applied in administration include certain branches of mathematics and certain quasi-mathematical systems of precise statement and reasoning.

The effect of these developments is to open new and enlarged opportunities for increasing administrative efficiency and rationality—both directly, through application of what is already known or already exists, and indirectly, through opening opportunities for new types of research on problems of organization and management. An example of direct application is operations research, which combines various specialized competences to analyze systems and solve problems. Developed in World War II as a technique for attacking certain problems in military strategy and logistics, operations research is now a speciality with an independent professional organization. Simulation is an example of the expanded potentialities for research. The new physical inventions, plus new or improved systems of logical thought, permit an unprecedented type and scale of simulation. Certain administrative situations (for example, decision-making problems) can be reproduced (simulated) in their essentials, thus helping to solve a perennial problem of social science: how to control and repeat experiments. [See OPERATIONS RESEARCH: SIMULATION.]

Many of these machines, logical systems, and what might be called social technologies, such as operations research, are now used in various branches of public administration; in some branches, such as military logistics and tax administration, they are widely used. On the whole, however, their development and use tends to center in business administration and schools of business administration. Perhaps, on the whole, they are more applicable (both presently and potentially) to business administration than to public administration. Nevertheless, increased knowledge about and use of them is on the agenda of Public Administration.

### The field outside the United States

As indicated above, wherever there is government there is public administration, but Public Administration—in the sense of a more or less autonomous discipline of alleged general applicability—was conceived and developed in the United States and is still strongly identified with its place of origin and greatest acceptance. This is not to assert, however, that in countries in which the American idea and content of Public Administration is not accepted the matter of training for pub-

lic administration is not taken seriously or that there may not be an application of science to the conduct of public administration.

Training for performance of some function in the governmental bureaucracy is a prime object of the system of public education of any country, in some cases nearly the sole object. In all advanced countries there is also some sort of special or differentiated training conceived as preparation for the positions of greatest responsibility and power in public administration—although other factors, such as social class, wealth, party membership, ideological purity, and religion, obviously are likely to enter in as selective criteria. In ancient China, as in modern Great Britain, for example, the regular and officially preferred training for high administrative positions was a humanistic education.

The most commonly accepted and approved training for high position, however, has been law (at least after and apart from training in military pursuits, for military organization is often the training ground for, and a source of supply of, governmental administrators for civil as well as military functions, especially in preindustrial societies). Presumably, this practice grows naturally from the fact that the making and enforcement of law are close to the heart of government; and in widely separated times and places people who know the law and have had their minds sharpened in its study have occupied the highest positions. In Anglo-American countries of the common law, though there is an important private quality in law, there has been a notable amount of this phenomenon; and in continental Europe, with its more immediate Roman background and its civil-law experience, where law is more strongly identified with the idea of the state, the predominant approach to preparation for high position is through study of law and jurisprudence, more particularly, that branch or aspect of the law identified as administrative law. This is not to say, again, that law and only law in a technical and formal sense is all that is taught and regarded as important. While in some situations it is substantially true that public administration is deemed to start and end with a reading and application of the book of detailed legal regulations, in some countries and in some courses of study law is deemed to be only a framework and is supplemented by other materials, including not only those from accepted social sciences (particularly economics), but perhaps from the "management science" around which Public Administration is formed.

Considerable discussion has taken place during the past generation between American and Conti-

mental students on the question whether it is more correct and fruitful to think and speak in terms of *an administrative science* or *the administrative sciences*. The American point of view has been that it is proper to speak of a *science of administration*, or a *science of public administration* (though the implications of these two expressions are rather different). Continental students, while willing to concede the importance of administration, have tended to see it rather as a place or process in which various disciplines or sciences closely interrelate.

To some extent, this controversy reflects a formal rather than a substantive difference: curriculums prescribed by the two camps for the training of public administrators might be very similar—for example, both might include sociology and economics, and perhaps administrative law and social psychology as well. At some point, however, there would likely come a parting of the ways, with the Europeans less likely than the Americans to include materials relating to “management science,” more insistent upon materials closely related to *their* national experiences and traditions, particularly in regard to law.

The communist countries share the Continental legal approach to the study and practice of public administration, though their heritage was in large part through the “second Rome” of Byzantium. This approach remains very strong, although the institutions and theories of communism make for important differences in public administration, and there are other influences, for example, the fact that some of the spirit and techniques of scientific management have been absorbed into communist administration.

### Trends, problems, and prospects

A review of the present situation in Public Administration should note some factors pertaining to the conduct and direction of public affairs in the United States that presumably will, or should, relate importantly to the development of the discipline—to the allocation of intellectual resources and the development of doctrines.

One factor is a continuing gradual increase in both the absolute and relative number of public employees. Approximately one in eight employed civilian persons is publicly employed, and while large numbers of these (publicly employed teachers, for example) are engaged in “line” functions, with little administrative responsibility, still the increase in numbers creates a growing problem in public administration: the increase in scale itself poses new problems for those who administer and teach

administrators, and the problems are intensified by an increasingly complex technology and an increasingly complex society. For example, to the extent that functions become more differentiated, and specialization and professionalization more advanced, in such fields as health, welfare, and education, the problem is posed of where and how the teaching of administration of such functions should take place. This is a problem of physical location and organizational arrangements; but it is also a problem in the evolution of science, or at least doctrine, posing the theoretical question whether administration is a “universal,” and the practical one whether it should be taught “in general” or as related to the content of the various fields. [See ADMINISTRATION; CIVIL SERVICE.]

Other factors concern changes and trends in public administration. In its origins Public Administration was concerned with certain problems of reform, conceptualized as problems in honesty and competence, and with such problems as departmentalization, executive control, and staff services, in which efficiency and economy were prime concerns. There has been much change and addition: human relations, communications, and so forth. But it is questionable whether recent and contemporary interests are as responsive as they should be to the content and direction of present governmental activities. Taking money as the measure, we might say that Public Administration pays relatively slight attention to two areas that represent more than half of public administration, namely, education and defense. Other important developments are inadequately represented in current professional attention; for example, the urban-metropolitan revolution in American life; the tremendous expansion of “government by contract and grant,” which has blurred the lines between the public and private sectors and between the federal and state spheres; and the rise of scientific research to the status of a pre-eminent national concern, a rise in which government at every level is involved in ways that lack even simple description and elementary understanding.

Public Administration since the 1940s obviously presents a spectacle of travail and transition, of controversy and confusion. The questions now are posed: Is Public Administration properly regarded as a discipline? Will it continue to be spelled with capital letters in connection with the organization of the intellectual life of the social sciences into curriculums, departments, professional societies, and so forth? The answer to the first of these questions is both a matter of definition and a matter of the unknowns of the future. The answer to the second



is very probably positive: Public Administration as an organizing concept will continue and probably grow in importance, although there is presently discernible some tendency to merge its activities and interests with related activities and interests under the broader term *public affairs*.

If *discipline* is defined very strictly as an intellectual enterprise with a body of consistent and agreed theory, then Public Administration is not a discipline and almost certainly will not become one. But few if any social sciences, or branches or disciplines thereof, fit this description. Indeed, few if any of the physical sciences do. If *discipline* is defined in terms of what has been called a core of unifying beliefs, then it is quite possible that in the future one of the now competing perspectives will achieve dominance, or a new synthesis will be achieved. Meanwhile, however, it should be emphasized both that there is no agreement on what constitutes a discipline and that intellectual progress does not await tidy definitions and agreements of opinion.

In regard to the question whether Public Administration will continue as an organizing concept, present trends indicate that its extent and importance, at least in the United States, will almost certainly increase. Concomitantly, there will certainly continue to be an increase in the complexity and difficulty of the problems of public administration in an increasingly complex society and a multiplication of new sciences and technologies bearing upon the processes and problems of administration, as well as continued advances in the old. There is likely to be more, not less, attention to the problems of organizing knowledge to bring it to bear on public administration as well as more, not less, training for and in public administration; and Public Administration will probably provide the organizing framework and the usual name for the central aspects of these enterprises.

We may achieve more clarity of concepts, more agreement among opposing schools of thought, or even a new synthesis; but it is clear that the value of Public Administration for designation of a focus of inquiry and activity, for the organization of research and instruction, does not depend on these eventualities. In fact, at the present stage of development, it may be better to regard Public Administration not as a single "thing," but as the customary and accepted collective term to designate a focus of interest and activity, in the way that the plural term *the administrative sciences* is used in Europe for a similar (but not identical) purpose. So used, it is analogous to *engineering* and *medicine* in indi-

cating a unifying focus for a wide array of sciences and technologies.

The recent history of Public Administration, and its present problems and opportunities, should be viewed from the perspective of the rejection in the 1940s of the division of politics and administration. That the political process reaches deeply into public administration and that the making of public policy is an important activity of public administration have since been all but universally recognized. But the response has been varied: there is still no general agreement on what conclusions follow, what the methodological implications are, what lines of inquiry are indicated as most appropriate and fruitful, what the implications are for programs of education and training. The discipline is presently trying to find its way through the tangle of issues. Many of the issues are posed in putative dichotomies and verbal antinomies: politics and administration, political science and Public Administration, science and art, pure science and practical science, fact and value, prescription and description, administrative science and administrative sciences, diversity and similarity, absolutism and relativism, universality and uniqueness, generality and concreteness. Whether in general or in any particular case the dichotomies are properly conceived, the antinomies properly opposed, is questionable, and they undoubtedly present a tangled skein to the student. In the last analysis, however, it must be universally conceded that important problems are involved.

The two most important immediate responses to the rejection of a separation between politics and administration were the application of the case method to the study of public administration and the reformulation presented in Simon's *Administrative Behavior*. These two responses were rather near the opposite ends of the spectrum of methodological possibilities.

The case approach has been motivated by a commitment to the objectives and methods of the social sciences, to be sure, but it has been shaped also by a considerable sensitivity to traditional concerns of the humanities and by a practical interest in pedagogy as against research. It sought the truth about administration by trying, dispassionately and painstakingly, to tell the story about what takes place in administration, in actual events, in context, and in all relevant dimensions. A large number of the abler students of the postwar generation in Public Administration were attracted to the case approach, and it continues to be of major importance.

The reformulation set forth in *Administrative Behavior* did not, paradoxically, attract many students in Public Administration—to whom it was presumably addressed—but it was and is highly regarded by students of business administration and apparently has been very influential in research in that field. Simply put, students in Public Administration were inclined to feel that Simon's work did not describe the world of public administration as they had experienced and observed it. To a generation emancipated from the politics-administration dichotomy, Simon's separation of fact and value seemed but another artificial division of the universe into two realms. Simon's commitment to a "hard" interpretation of social science, his philosophical and methodological program, were unappealing. However, younger students, more familiar with and affected by behavioralism, probably find the formulations more meaningful and appealing, and it is possible that the work thus may have a delayed, direct impact on the discipline, in addition to the indirect influence that it has already exercised.

The most significant development in Public Administration, currently engaging the attention and energies of a large number of students both young and mature, is the focusing of attention on comparative public administration and the related problems of "development administration." This attention grows out of personal experience, is related importantly to world-wide developments that are likely to continue, and is addressed to a wide spectrum of interests from concrete policy questions to the abstractions of the pure social sciences. [See TECHNICAL ASSISTANCE; see also FOREIGN AID.]

For the most part, it is the younger students in Public Administration who are active in the comparative movement, and certainly it is they who are chiefly interested in the theoretical-scientific questions. For the most part and in a general sense these younger students are behaviorally oriented; they are knowledgeable about and have a concern with the central problems of the social sciences. But the paradox noted above is further exemplified: They are not especially attracted by the formulations and interests of Simon, but find their inspiration, models, and techniques in other parts of the contemporary social sciences, most notably in the companion movement in political science—comparative politics—and in sociology.

From the point of view of the scientific study of public administration, the course of future developments would seem to rest in large measure on the respective futures of the comparative public ad-

ministration movement and the administrative science point of view. By the latter is meant the outlook and research interests (including but not limited to those of Simon) that are identified in a general way with business administration; in a general way they grow out of the old scientific-management movement as modified by the Hawthorne studies; are strongly oriented historically toward efficiency and currently toward the related, but broader, "rationality" (or the even broader decision-making schema); make considerable use of mathematics; include but are not limited to the new technologies and techniques discussed above; and aspire to a science of administration in every essential way as scientific as current physical science. Administrative science as an approach or movement is closely joined with, but still distinguishable from, the theory-of-organizations movement. Both movements—particularly the theory of organizations—are sometimes concerned with "comparativeness" in some aspect, but their concern is largely with intracultural, not intercultural comparison.

In the broadest sense, the future of Public Administration is engaged with the future of political science, on the one hand, and with administrative science, on the other. It has, from its beginnings, represented a joining of certain interests of political science with the "management" movement; and it still does, granted all the additional factors, the new developments, the broadened spectra. What meaning and importance will be given to the Public in Public Administration, whether it will evaporate or remain significant, perhaps even become more significant, will depend in large part on what developments take place in political science and in the social sciences—indeed in contemporary thought—as a whole.

DWIGHT WALDO

[Directly related are the entries ADMINISTRATION; CIVIL SERVICE; DECISION MAKING; INTERNATIONAL ORGANIZATION, article on ADMINISTRATION; POWER; PUBLIC POLICY. Other relevant material may be found in BUREAUCRACY; GOVERNMENT; LAW; ORGANIZATIONS; POLITICAL SCIENCE.]

#### BIBLIOGRAPHY

- APPLEBY, PAUL H. 1952 *Morality and Administration in Democratic Government*. Baton Rouge: Louisiana State Univ. Press.
- DAHL, R. A. 1947 The Science of Public Administration: Three Problems. *Public Administration Review* 7:1-11. → A brief, incisive critique of "orthodoxy."
- DAVY, THOMAS J. 1962 Public Administration as a Field of Study in the United States. *International Review of Administrative Sciences* 28:63-78. → An excellent re-



- view of trends, in which the major contemporary orientations are treated under four headings: managerial, political, psychological, and sociological.
- DE GRAZIA, ALFRED 1960 *The Science and Values of Administration*. *Administrative Science Quarterly* 5:363-397, 557-582. → Observations and conclusions by a leading behaviorist, presenting an "action" schema for the understanding of administration.
- GAUS, JOHN M.; WHITE, LEONARD D.; and DIMOCK, M. E. 1936 *The Frontiers of Public Administration*. Univ. of Chicago Press. → An outstanding work of the period, foreshadowing later developments.
- GOODNOW, FRANK J. 1900 *Politics and Administration: A Study in Government*. New York: Macmillan. → A classic work in American political science and Public Administration.
- GULICK, LUTHER H.; and URWICK, LYNDALL (editors) (1937) 1954 *Papers on the Science of Administration*. New York: Institute of Public Administration. → Regarded as the high-water mark of "orthodoxy"; much of it still interesting and relevant to present concerns.
- HEADY, FERREL 1966 *Public Administration: A Comparative Perspective*. Englewood Cliffs, N.J.: Prentice-Hall. → Discusses the significance of comparative studies for Public Administration and relates public administration to political systems.
- LANDAU, MARTIN 1962 *The Concept of Decision-making in the "Field" of Public Administration*. Pages 1-28 in Sidney Malick and Edward H. Van Ness (editors), *Concepts and Issues in Administrative Behavior*. Englewood Cliffs, N.J.: Prentice-Hall. → An incisive essay, critical of present intellectual orientations and urging a sharper scientific approach centered on decision making.
- LONG, NORTON E. 1962 *The Polity*. Chicago: Rand McNally.
- MOLITOR, ANDRÉ 1959 *The University Teaching of Social Sciences: Public Administration*. Paris: UNESCO. → A valuable world-wide review, discussing among other matters the issue of "administrative science" versus "administrative sciences."
- MOLITOR, ANDRÉ 1961 *Public Administration Towards the Future*. *International Review of Administrative Sciences* 27:375-384. → Observations and estimates on present and future developments, distinguishing between developed and developing countries.
- MOSHER, FREDERICK C. 1956 *Research in Public Administration: Some Notes and Suggestions*. *Public Administration Review* 16:169-178. → Contains valuable reflections on the status and problems of Public Administration.
- PITTSBURGH, UNIVERSITY OF, ADMINISTRATIVE SCIENCE CENTER 1959 *Comparative Studies in Administration*. Univ. of Pittsburgh Press. → See especially the Foreword and the Introduction for a statement of the administrative science point of view.
- REDFORD, EMMETTE S. 1958 *Ideal and Practice in Public Administration*. University: Univ. of Alabama Press.
- ROBSON, WILLIAM A. 1961 *The Present State of Teaching and Research in Public Administration*. *Public Administration* (London) 39:217-222. → A British view of current trends, rather negative regarding American theories and hopes. The journal is the main source of information on the British approach to the study of public administration.
- ROETHLISBERGER, FRITZ J.; and DICKSON, WILLIAM J. 1939 *Management and the Worker: An Account of a Research Program Conducted by the Western Electric Company, Hawthorne Works, Chicago*. Cambridge, Mass.: Harvard Univ. Press. → A paperback edition was published in 1964 by Wiley.
- SAYRE, WALLACE S. 1958 *Premises of Public Administration: Past and Emerging*. *Public Administration Review* 18:102-105.
- SIMON, HERBERT A. (1947) 1961 *Administrative Behavior: A Study of Decision-making Processes in Administrative Organization*. 2d ed. New York: Macmillan. → A seminal work, offering both a critique of accepted doctrine and a program for reconstruction.
- WALDO, DWIGHT 1948 *The Administrative State: A Study of the Political Theory of American Public Administration*. New York: Ronald Press. → A thorough historical review of the rise of Public Administration, containing also a critique of doctrines.
- WALDO, DWIGHT 1955 *The Study of Public Administration*. New York: Random House; Garden City, N.Y.: Doubleday.
- WALDO, DWIGHT 1965 *The Administrative State Revisited*. *Public Administration Review* 25, no. 1:5-30. → Discusses doctrinal developments and problems in Public Administration since the publication of *The Administrative State* in 1948.
- WHITE, LEONARD D. (1926) 1955 *Introduction to the Study of Public Administration*. 4th ed. New York: Macmillan.
- WILLOUGHBY, WILLIAM F. 1927 *Principles of Public Administration: With Special Reference to the National and State Governments of the United States*. Baltimore: Johns Hopkins Press.
- WILSON, WOODROW (1887) 1953 *The Study of Administration*. Pages 65-75 in Dwight Waldo (editor), *Ideas and Issues in Public Administration*. New York: McGraw-Hill. → An interesting and significant historical document in Public Administration.

## PUBLIC ASSISTANCE

See MEDICAL CARE; PLANNING, SOCIAL, article on WELFARE PLANNING; SOCIAL PROBLEMS; SOCIAL WORK; WELFARE STATE.

## PUBLIC DEBT

See DEBT, PUBLIC.

## PUBLIC ENTERPRISE

See NATIONALIZATION.

## PUBLIC EXPENDITURES

In dealing with the theory of public expenditures, two approaches may be taken. We may take a normative view and explore the role that public expenditures should play in an efficiently functioning economy. Or we may examine the sociology or politics of fiscal behavior, explaining the forces that determine actual expenditure policy in the

prevailing historical and institutional context. This essay is directed largely to the normative question, attention being given in turn to three major aspects of expenditure policy—the allocation, distribution, and stabilization functions. In conclusion, brief consideration is given to fiscal sociology or politics.

### Allocation function

The central issue in the theory of public expenditures, and, indeed, in the theory of public finance, is how to determine the proper level and pattern of public services. Putting it differently, the question is how available resources should be divided between the satisfaction of "private" and of "social" wants. Looked at as an economic problem, this immediately poses a second issue. If resources are to be used for public services or for the satisfaction of social wants, which private services or satisfactions are to be forgone? As public services must be justified in terms of their opportunity cost, the theory of expenditure-making cannot be separated from the theory of expenditure-financing. Basically, the theory of public expenditures and the theory of taxation are but different sides of the same coin.

Attempts at dealing with this problem have a long and frustrating history. Adam Smith, in Book v of *The Wealth of Nations*, squarely addressed himself to both its aspects. Certain services, so he argued, must be provided by the state, including the upkeep of the sovereign, defense, some phases of education, and certain public works that require too much capital or are too distant in their payoff to be undertaken privately. The finance in turn should be provided through taxes levied so that all will contribute "as nearly as possible in proportion to their respective abilities; that is, in proportion to the revenue which they respectively enjoy under the protection of the state" (Smith 1776, vol. 2, p. 307 in 1910 edition). Ingeniously embedded in a single formula, this rule contains the roots of both the "ability to pay" and the "benefit" theory, the two approaches that were to provide the major strands of future discussion.

The ability-to-pay doctrine became the dominant view among British writers. It was reformulated by J. S. Mill in 1848 in terms of equality of sacrifice, translated by Edgeworth in 1897 into a requirement of least total sacrifice, and refined by Pigou in 1928 into a choice between equality of absolute, proportional, or marginal sacrifice. As a doctrine, it was attractive both to social reformers, who looked upon it as an instrument of income equalization, and to conservatives, who shied away from the more positive view of public expenditures embedded in the benefit approach.

Notwithstanding its wide and sustained popularity, the ability-to-pay approach has severe shortcomings. By its very nature, it deals with taxation only and fails to establish a link between the tax and the expenditure sides of the budget problem. At best, it can determine how a given tax total is to be distributed among various taxpayers, independently of the budget composition. And even this achievement, unsatisfactory as it is because the crucial problem of opportunity cost is neglected, has become increasingly dubious.

The determination of equal sacrifices involves direct interpersonal utility comparisons. Such comparisons, which were the foundation of welfare economics up to the 1930s, were discarded thereafter as lacking operational content. The new formulation, by which there is a gain in total welfare only if A's position is improved without worsening B's, overcomes this difficulty but is also much more limited in application. In particular, it is inherently useless in dealing with distributional problems. If this new formulation is accepted, the ability-to-pay concept merely survives as a reflection of what is considered a socially desirable distribution of income; and once this view is taken, there is no reason why the scope of redistribution should be limited to the financing of public expenditures on goods and services, rather than including transfers as well. [See WELFARE ECONOMICS.]

The benefit approach, from the beginning, had the advantage of linking the revenue and expenditure sides of the budget problem. The question, however, is whether these individual gains can be measured, since otherwise the benefit principle lacks operational meaning. A first answer was given by Smith's hypothesis that the essential service of government is protection and that the value of protection received may be measured in terms of income. This protection version of the benefit approach was subsequently adopted by many Continental writers. In many instances, it provided a vehicle for imposing severe limits on the scope of budget activity, and through most of the nineteenth century the benefit discussion fell far short of its inherent possibilities.

These possibilities, indeed, did not become apparent until the close of the century when the then new economics of marginal utility calculus came to be applied to the problems of public finance.

The participants in this debate included Sax in 1887, Pantaleoni in 1883, Mazzolo in 1890, Viti de Marco in 1928, and many others; but the most important contribution, from the modern point of view, was made by Wicksell (1896). This was followed by the subsequent elaboration of his work



by Erik Lindahl in 1910 and 1919 (see Musgrave & Peacock 1958 for excerpts from these writers).

Wicksell began with the assumption that there exists a "proper" distribution of income, holding that otherwise there can be no just determination of tax shares. Voters are then called upon to vote on various budget proposals, including a wide range of tax-expenditure combinations. For an expenditure plan to be passed, it should be paired with a financing plan so that it receives unanimous acceptance or, at least, acceptance by a substantial majority vote. With remarkable intuition, Wicksell thus advanced a solution under which the budget plan results in a gain in welfare, even as judged by the severe standards of modern welfare economics.

It is important to note that the approach of these writers takes a strictly individualistic view of social wants. Social wants, along with private wants, are included in the preference system of the individual, and social no less than private wants should be satisfied in accordance with the sovereign choice of consumers. But if there is no difference in the psychology of private and of social wants, why need the distinction be drawn? And if private wants are satisfied through the market, why cannot the same be done for social wants, i.e., why is a public budget needed?

The answer is that there is an important difference; but it is technical, not psychological. In the extreme case of a purely social want—or a purely social good provided to satisfy such a want—we are dealing with a good whose benefits are altogether "external," that is, they are not limited to any particular consumer. This has two important implications that explain why social wants cannot be satisfied through the market mechanism but must be provided for through the budget process.

First of all, note that the nature of social goods is such that all consumers are confronted with more or less the same quantity supplied. The benefits of national defense accrue to all citizens, and smog removal helps everyone in the area. While individual consumers may purchase private goods in whatever particular amounts they find suitable, all must purchase more or less the same amount of public goods. Assuming two taxpaying consumers A and B, A's demand schedule together with the total cost schedule for public services will determine the price at which the service is available to B, and B's demand schedule will, in turn, determine the price to A. According to Lindahl, the optimal supply of public services and the determination of proper cost shares is arrived at where both demand schedules intersect. This is the point

at which both A and B agree on the amount of public services which they prefer, given the underlying cost distribution. Following Bowen's version (1948) of a similar view, total demand is determined by adding individual demand schedules vertically (as distinct from horizontally for the case of private goods). Output is determined by the intersection of the total demand and supply schedules; and the offer-price indicated by each demand schedule indicates the proper cost shares of the various demands. Thus the determination of expenditure level and tax-price is again seen as proceeding analogously to that of market supply and market price for private goods.

Restating the argument in terms of general welfare theory, Samuelson (1954) defined a social good as one the consumption of which by A does not reduce its availability for consumption by B. Because of this "joint consumption" feature, the condition of Pareto optimality for the social good is shown to differ from that for the private good. Assuming preferences to be revealed, a planner may determine a set of Pareto-optimal solutions, involving various divisions of total output between social and private goods, and distributions of private goods between A and B. While the Lindahl solution gives one such point on the utility frontier, it is not the only one. The choice among the points on the frontier becomes a distributional issue, to be decided in the context of a general social welfare function. At the same time, the Lindahl solution has certain unique characteristics (analogous treatment of social and private goods from the point of view of the consumer) that render it especially attractive.

Second, note that this formulation, although complex enough, greatly oversimplifies the problem. It assumes that true consumer preferences are, or will be, revealed readily through the voting process. This assumption is unrealistic. In the case of private wants, a consumer must pay at the counter if he wishes to purchase the requisite private good; and in competing with other consumers for the supply of such goods, he is forced to reveal his preferences. The market, as it were, functions as an auction system. In the case of social wants the situation differs. A consumer may enjoy the supply of social goods without having to qualify by paying at a counter. This being the case, he is not induced to bid for such services. But cannot this difficulty be overcome by having the budget office issue a bill and enforce payment? Unfortunately, this is no solution. The real difficulty lies not in enforcing payment of taxes once they are properly deter-

mined, but rather in determining what and how much should be spent, and how much particular individuals should be called upon to contribute.

In order to induce revelation of consumer preferences, a political voting process is needed, parallel in function to the private voting (bidding) process in the market. The complication arises because—unlike the case of market bidding—voters (consumers) cannot be simply assumed to record their true preferences. Since any particular individual cannot be excluded from the enjoyment of social goods, he will be tempted to understate his true preferences. If the number of voters (consumers) is large, the effect of any one contribution on the total supply of social goods will be negligible. If the number of voters is small, collusion and imperfections (analogous to interference with competition in the private market) will develop. Strategy will enter, and the result will differ from that which would be obtained under the above solution in which true preferences were revealed.

A dilemma arises because the type of voting mechanism (for example, point voting) best designed to give efficient results in the absence of strategy also lends itself most to abuse by strategy; whereas a system less sensitive to strategy (for example, a crude majority vote) does not permit the voter to record the degrees of intensity of his preferences required for an efficient solution. It is thus left to the "art" of politics to establish a workable system of preference determination and tax assessment, where a constructive compromise is reached between these two sides of the problem. Indeed, this is one of the central tasks of government in a democratic society.

The question arises whether the specific characteristic of the social good lies in the "joint consumption" feature or in "non-excludability." It appears that the former is more basic because application of exclusion would be undesirable for the social good even if it were possible. Since B's consumption does not interfere with A's, it would be inefficient to exclude him.

We have seen that, for reasons inherent in the technical nature of social goods, social wants must be provided for through the public budget. To avoid misunderstanding, the expression "provided for" needs brief interpretation. Budgetary provision for social goods does mean that the supply of such goods and the assignment of their cost must be determined through the tax-expenditure process of government. It does not mean that social goods (i.e., goods supplied to meet social wants) must be produced under public management. Highways and

planes may be provided for publicly, yet be purchased from private construction or aviation companies. (The market can function in supplying such goods once government appears as a demander. The problem is in establishing government demand.) Similarly, the government may produce private goods and distribute them through sale on the market, as illustrated by the traditional tobacco monopoly. The issue of public versus private production is the essential issue of capitalism versus socialism, but not of fiscal theory. It must not be confused with the fiscal problem of allocating resources between the satisfaction of social wants and the satisfaction of private wants. Indeed, this fiscal problem exists in the confines of a socialist as well as a capitalist system.

**Complications.** We now turn to some major qualifications or objections that may be raised regarding this statement of expenditure theory.

(1) The argument, so far, has been presented as if there existed a clear-cut distinction between two types of goods—some that are purely private (benefits accruing to the owner only) and others that are purely social (benefits being altogether external). Actually, most goods involve both characteristics to varying degrees. It is hard to find budget items, other than defense, that do not involve private-want components. Education expenditures, in particular, are of the mixed type. Similarly, many privately purchased goods involve social-want aspects, be they positive (shade trees next door) or negative (smoke nuisance).

This qualification is quite in order, but it is not in conflict with the above approach. Our argument may be readily adapted to the case of mixed goods, so that budgetary provision is made for that part of the cost reflecting benefits of the social-want type, while private payment is made for that part reflecting benefits of the private-want type. The theory of taxation is thus converted into a more general theory of subsidy, the case of the 100 per cent social good (and the 100 per cent tax finance) being reduced to that of a limiting case only, matched at the other end by the 100 per cent private good (and 0 per cent tax finance).

(2) Our formulation bases the distinction between social and private wants on the technical (externality) characteristics of social goods. As far as the psychology of wants is concerned, both private and social wants are considered homogeneous parts of the individual's preference system, and both are to be satisfied in line with the principle of consumer sovereignty, i.e., in line with the preferences of the individual consumer. In contrast to



this view, some writers (especially in the German literature) have argued that social wants should not be interpreted in terms of individual preferences but should be considered as deriving from collective or community needs. The question is how and by whom these community needs are experienced and determined. If an "organic" view of the state is taken, it remains to be shown how the organism reveals its choice. If the expenditure choice and cost allocation are made by a set authority (tribal chief, king, dictator), the problems here posed are much simplified, but the resulting theory of the public sector is hardly satisfactory in a democratic setting. [See CONSUMER SOVEREIGNTY.]

An alternative formulation of the criticism, more acceptable to the democratic setting (Colm 1955), may be that the choice should be made through the democratic process, but that the individual (while participating therein) should be taken to be acting in the social interest rather than in his personal interest. This makes sense where matters of national concern, such as defense policy, are at stake. In such cases the social interest will be reflected in individual preferences. But there are many other instances (e.g., highways for private or commercial use) where the good is neither more nor less trivial than is the ordinary type of private good and where the ordinary motivation of consumer choice is altogether appropriate. The mere fact that externalities are attached to certain goods complicates the process of resource allocation; but it establishes no presumption that such goods serve more noble purposes or that basic preferences for such goods must be set on a different moral basis.

(3) Nevertheless there are instances, even in the most democratic of societies, where individual consumer choice is overruled. On the negative side, minors are not permitted to buy guns, and the availability of drugs to adults is restricted as well. On the positive side, certain services, the supply of which could be made subject to purchase payments (for example, education), are supplied free of charge or subsidized because individual outlays thereon are considered insufficient. Or, redistributive measures are taken through payments in kind (subsidized low-cost housing, for example) rather than through cash transfers.

Such infringements on consumer choice cannot be banned altogether from a democratic society. Constructively used, they can be a helpful aid in the social learning process. This part of the budget process—which I have referred to in another context (Musgrave 1959, p. 13) as the satisfaction of "merit wants"—cannot readily be fitted into the

above theory of expenditure determination. While this is a shortcoming, merit wants are the exception rather than the rule; and their existence does not invalidate the usefulness of our model, which covers the major and more manageable part of the budget problem.

(4) The benefits to be derived from particular expenditure projects are frequently difficult to assess, so that the voter (or his representative) needs expert guidance in making the appropriate choices. It is here that the role of the civil servant and budget expert is of great importance.

Recently, this problem has received special attention with regard to outlays on public investment. Any given public investment should be undertaken only if the rate of return exceeds or equals that of an alternative public or private investment. The problem then is how to measure the rate of return on public investment. This involves a number of important issues. One is the determination of the appropriate rate of discounts, the issue on which much of the recent discussion has focused. Another less explored but perhaps more important matter is the development of techniques by which secondary benefits resulting from a particular project may be measured. Finally, there is the problem of developing rules of thumb (e.g., various types of cost-benefit ratios) that can be used for guidance in actual policy making. [See INVESTMENT, *article on* THE INVESTMENT DECISION; WATER RESOURCES.]

(5) In the modern democratic society, budgetary matters are not decided by referendum but in the legislature. This does not mean that the decision is made without relation to the preferences of the individual voter. The legislator acts by representing his constituents and is elected by exhibiting a platform more or less reflecting their preference patterns. By engaging in coalition with other legislators, he serves to establish bundles of issues on which a consensus can be reached. All this is an important part of preference determination and may be looked upon as a means of expediting, rather than displacing, individual choice. Of course the party mechanism, in some of its manifestations, may also act as an interfering factor.

(6) The assumption that all citizens are confronted with the very same supply of social goods needs to be qualified, especially with regard to regional differentials. The regional aspects of expenditure theory (for example, regional differences in benefit incidence or spillover of benefits between regions with different government) cannot be dealt with here, nor can the qualifications needed as the model of a unitary government (implicit in our

discussion) is replaced by one of fiscal federalism. [See LOCAL FINANCE.]

### Distributional function

So far, we have dealt with budgetary provision for goods and services designed to satisfy social wants. This "allocation" aspect of the public household, although perhaps the most basic function of the revenue-expenditure process of government, is by no means the only one. Distributional adjustments must also be considered.

If tenets of modern welfare economics are accepted, economics, as previously noted, has nothing to say on the basic issue of income distribution. While the economic analyst may explore the consequences (with regard to the level of output, growth, and other factors) of various changes in distribution, he cannot compare the merits of alternative distributions of a given output. This could be done only if interpersonal utility comparison is admitted and an operational procedure of comparison could be devised.

However this may be, distribution does present a policy issue. This is obvious in the socialist setting, where the return to capital accrues to the state and wages paid need not equal the return to labor as a planning cost. But even in the most capitalistic of countries, distribution is not left entirely to the ownership of factors (labor as well as capital and natural resources) and the market system of factor pricing. Some degree of intervention is held necessary, if only to provide for the indigent. Beyond this, society may consider the unadjusted state of distribution to be less or more equal than is held desirable and choose to make the necessary adjustments. The necessity for some distributional adjustments is thus generally accepted, even though the desired degree of adjustment is highly controversial. [See POVERTY.]

The most direct and efficient tool for carrying out such adjustments as are desired is provided through the tax-transfer process. Use of the tax-transfer instrument is superior to interference with factor pricing, which gives rise to inefficient resource use in the private sector of the economy. It is superior as well to interference with product prices (subsidies) or redistribution in kind (e.g., distribution of free goods rather than cash), since transfers do not interfere with the option of consumer choice. Such at least is the case unless redistributive objectives coincide explicitly with situations where interference with consumer choice is held desirable in order to satisfy merit wants.

Reasoning along these lines, an orderly budget system would provide for two subbudgets. There

would be a tax-purchase budget, designed to provide for goods and services needed to satisfy social wants. Also, there would be a tax-transfer budget designed to make such distributional adjustments as are held desirable. Assuming correction in the direction of greater equality, the latter budget would involve progressive taxes combined with regressive transfers. The distribution of taxes in the former budget would depend on the income and price elasticities of demand for public services. The resulting distribution of taxes in the allocation budget could be proportional, progressive, or regressive, depending on the circumstances of the case.

Given these two subbudgets, one may then think of a combined budget, where the two sets of taxes are combined and netted out against transfers. Assuming the tax distribution under the allocation budget to be proportional to income, with tax transfers toward greater equality in the distribution budget, the net budget would show a progressive structure, with transfers at the lower end of the income scale. In this way, the tax-transfer pattern in the net budget may be interpreted in a rational manner, and the vague and unsatisfactory notion of "ability to pay" is replaced by a more clear-cut system of budget determination.

All this is but a way of formalizing the Wicksellian principle of separating the issue of budget policy for redistribution from that of budget policy for the satisfaction of social wants. This separation has great merit for an orderly view of budget policy. At the same time, it must be granted that complete separation is not possible. Merit-want objectives may coincide with distributional purposes, the assumption being (presumably) that low-income consumers are most in need of consumption guidance. Moreover, we have seen that the very nature of social wants does not assure a single best solution to the budget problem and, in a sense, requires a distributional decision to choose among the points on the utility frontier. How may this be squared with the Wicksellian dictum that a "just" allocation of the cost of public services requires a "just" distribution of income?

As has been pointed out by modern welfare economists, a proper state of distribution must be defined in welfare terms. Assume that there are private goods only. A given state of income distribution may then lead to different welfare distributions (different points on the utility frontier) depending on which pricing rule is used. Since there is more than one efficient rule, a meaningful concept of "proper" income distribution must be linked with a specific pricing rule. In the case of private goods,



this is supplied conveniently by uniform price-marginal cost pricing in a competitive market. An analogous argument holds if the allocation and distribution problems are separated in the case of social goods. If a given income distribution is said to be proper, it must be specified as such on the assumption that a given pricing rule is followed with regard to social goods. One such rule is given by the Lindahl solution, where the tax price paid by each consumer is such that his marginal rate of substitution of public for private goods must equal the ratio of public goods prices payable by him to private goods prices.

### Stabilization and growth

Having considered the allocation and distribution function, it remains to note the stabilization function of budget policy. In a decentralized market system, there is no assurance that the level of aggregate demand may not at times be excessive (inducing inflation) or deficient (inducing unemployment). To provide the necessary corrective, monetary and/or fiscal measures may be needed. Expansionary fiscal action may be taken in various ways, including expenditure increase, tax reduction, and balanced budget expansion.

In the context of a general theory of expenditure determination, the stabilization function is to be performed in such a way as not to interfere with the previously examined allocation and distribution functions. This suggests that expansionary action takes the form of proportional tax reduction (or transfer increase), while restrictive actions take the form of proportional tax increase (or transfer reduction). Putting it differently, the level of government expenditures on goods and services should be so set as to provide for proper allocation of resources under the condition of full resource use, leaving it to the tax-transfer mechanism to secure full employment; and the use of this mechanism should be such that it will not affect the "proper" state of distribution.

It does not follow, however, that the level of goods and service expenditures should never be varied over the cycle. The very factors that make for changes in the level of private expenditures (for example, changes in investment opportunity) may affect (in the opposite direction) the demand for public services; problems of unemployment may be regionally confined and require regionally focused action; public jobs may be more desirable socially than the public dole, and so forth. Thus, countercyclical expenditure policy cannot be ruled out, even though there is a *prima facie* argument for tax-transfer adjustments.

Recent emphasis in fiscal policy discussion has been upon the growth, rather than the countercyclical, aspects of stabilization policy. This has important bearing on the allocation function of expenditure policy. Given the level of resource utilization, increased growth requires a higher level of capital formation relative to consumption. This means (a) emphasis on public investment, relative to public consumption, in the allocation function of budget policy; and (b) a mixture of fiscal and monetary policy that is inductive to a high rate of private capital formation. Under favorable conditions (a high level of potential investment in the private sector), this may be secured by combining easy availability of credit with a tight (high-tax) budget policy. Under less favorable conditions, expansion of public investment and special (tax or transfer) incentives to private investment may be needed.

Following the procedure suggested for the distribution function, we may now visualize a third budget, addressed to the stabilization function. In securing high employment and price-level stability, and doing so without interference with the other purposes of budget policy, the stabilization budget calls for proportional taxes when restriction of private spending is needed and for transfers when expansion is required. Whereas the distribution budget is always balanced (taxes on A being used to make transfers to B) and the allocation budget is balanced over time (consumption expenditures being tax financed and capital expenditures being amortized over the life of the asset), the stabilization budget by its very nature is either overbalanced or underbalanced.

Accordingly, we may now think of the over-all budget as reflecting a consolidated version of the three subbudgets, reflecting the net patterns of expenditure and tax payments, as well as the state of net balance. While this is the budget that, in effect, comes to be administered, disaggregation is needed to understand the policy issues involved.

### Fiscal politics

The preceding discussion was addressed to the question of how to secure efficient conduct of the public sector in a democratic society. An alternative approach to the theory of public expenditures is to provide an explanation of the actual conduct of expenditure policy.

This approach, which will be noted but briefly, may be referred to as fiscal sociology or fiscal politics. Karl Marx, in the *Communist Manifesto*, held that capitalism might be undermined by progressive taxation, just as it would be weakened (and

this should be added to place his judgment in proper perspective) by popular education. Adolph Wagner predicted a rising ratio of public expenditures to Gross National Product as a law of social and political development. Wicksell foresaw the possibility that a fiscal mechanism guided by majority voting could lead to the exploitation of the rich by the poor, thereby reversing the earlier concern of social reformers. Rudolf Goldscheid presented a sweeping theory of budget behavior, cast in the context of a Marxist view of history. The wealthy state of feudal society, so he argued, gave way to the impoverished modern state, precisely as the ruling class relinquished control over the state to the people (see Musgrave & Peacock 1958 for excerpts from these writers). Joseph Schumpeter (1918), taking off from Goldscheid, re-examined the change of fiscal forms in the transition from feudalism to capitalism, and so forth.

The present writer has stressed the importance of political interaction between the various functions of budget policy as an explanatory as well as distorting factor in budgetary decision making. Emphasis has been placed by Günter Schmolders (1955) on the psychological basis of taxpayer behavior, and various attempts have been made—including those by Downs (1957), Buchanan and Tullock (1962), and Black (1958)—to explore the consequences of party behavior and majority voting.

Some have argued that the political process tends to result in an excessive level of public expenditures because voters tend to underestimate the cost of public services, expecting it to be paid by the next fellow. Others have held the opposite, whether it is, as suggested by Galbraith (1958), because the need for private services is overrated in response to advertising, because the political mechanism is biased toward inaction rather than action, or because benefits from public services are less urgently required. Moreover, the intrusion of distribution or stabilization considerations into the allocation budget may lead to a distortion of the expenditure level, rendering it unduly high when expansionary action is needed or when the ruling majority considers budget expansion as a means toward increased redistributive gains, and vice versa under opposite conditions. Finally, it has been suggested by Peacock and Wiseman (1961) that the growth of public expenditures is related to the tendency for increases in tax rates during crisis periods such as war to be nonreversible, thus periodically sucking up the level of public outlays for civilian purposes. None of these explanations is altogether convincing; but they suggest that fiscal

sociology, or fiscal politics, is a fascinating, if as yet largely unexplored, area of research.

RICHARD A. MUSGRAVE

[See also BUDGETING, article on GOVERNMENT BUDGETING; FISCAL POLICY; TAXATION; WELFARE ECONOMICS; and the biographies of PIGOU and WICKSELL.]

#### BIBLIOGRAPHY

- ARROW, KENNETH J. 1951 *Social Choice and Individual Values*. New York: Wiley.
- BLACK, DUNCAN 1958 *The Theory of Committees and Elections*. Cambridge Univ. Press.
- BOWEN, HOWARD R. 1948 *Toward Social Economy*. New York: Rinehart.
- BUCHANAN, JAMES M. 1960 *Fiscal Theory and Political Economy: Selected Essays*. Chapel Hill: Univ. of North Carolina Press.
- BUCHANAN, JAMES M.; and TULLOCK, GORDON 1962 *The Calculus of Consent: Logical Foundations of Constitutional Democracy*. Ann Arbor: Univ. of Michigan Press.
- COLM, GERHARD 1955 *Essays in Public Finance and Fiscal Policy*. Oxford Univ. Press.
- DOWNES, ANTHONY 1957 *An Economic Theory of Democracy*. New York: Harper.
- ECKSTEIN, OTTO 1961 A Survey of the Theory of Public Expenditure Criteria. Pages 439-504 in Universities-National Bureau Committee for Economic Research, *Public Finances: Needs, Sources, and Utilization*. National Bureau of Economic Research, Special Conference Series, No. 12. Princeton Univ. Press. → This survey includes extensive references to the literature.
- GALBRAITH, JOHN KENNETH 1958 *The Affluent Society*. Boston: Houghton Mifflin.
- JOHANSEN, L. 1963 Some Notes on the Lindahl Theory of Determination of Public Expenditures. *International Economic Review* 4:346-358.
- MARGOLIS, JULIUS 1955 A Comment on the Pure Theory of Public Expenditures. *Review of Economics and Statistics* 32:347-349.
- MARGOLIS, JULIUS 1957 Secondary Benefits, External Economies and the Justification of Public Investment. *Review of Economics and Statistics* 39:284-291.
- MUSGRAVE, RICHARD A. 1959 *The Theory of Public Finance: A Study in Public Economy*. New York: McGraw-Hill. → See especially Parts 1 and 2.
- MUSGRAVE, RICHARD A.; and PEACOCK, ALAN T. (editors) 1958 *Classics in the Theory of Public Finance*. London and New York: Macmillan. → Contains excerpts from the more important European writers.
- PEACOCK, ALAN T.; and WISEMAN, JACK 1961 *The Growth of Public Expenditure in the United Kingdom*. National Bureau of Economic Research, General Series, No. 72. Princeton Univ. Press.
- SAMUELSON, PAUL A. 1954 The Pure Theory of Public Expenditures. *Review of Economics and Statistics* 36: 387-389.
- SAMUELSON, PAUL A. 1955 Diagrammatic Exposition of a Theory of Public Expenditures. *Review of Economics and Statistics* 32:350-356.
- SCHMOLDERS, GÜNTER 1955 *Finanzpolitik*. Berlin: Springer.
- SCHUMPETER, JOSEPH A. (1918) 1954 The Crisis of the Tax State. *International Economic Papers* 4:5-38. → First published as *Die Krise des Steuerstaats*.



SMITH, ADAM (1776) 1952 *An Inquiry Into the Nature and Causes of the Wealth of Nations*. Great Books of the Western World, Vol. 39. Chicago: Encyclopædia Britannica. → A two-volume paperback edition was published in 1963 by Irwin.

WICKSELL, KNUT 1896 *Finanztheoretische Untersuchungen nebst Darstellung und Kritik des Steuerwesens Schwedens*. Jena (Germany): Fischer.

## PUBLIC FINANCE

For an overview of the economic aspects of government receipts and expenditures, see BUDGETING, article on GOVERNMENT BUDGETING; FISCAL POLICY; LOCAL FINANCE. The most general treatment of government expenditures is PUBLIC EXPENDITURES; more specialized aspects are treated in SUBSIDIES; CREDIT, article on GOVERNMENT CREDIT. For receipts, see TAXATION; for government borrowing, see DEBT, PUBLIC.

## PUBLIC HEALTH

As far back in time as one can ascertain, men have lived in the organized groups that we call communities and have had to deal in one way or another with health problems arising from their biological attributes and needs and from their social circumstances. The major areas of concern have been the provision of food and water in sufficient supply and of acceptable quality, the control of the physical environment, the prevention or control of epidemic and endemic diseases, the provision of health care, and the relief of physical and social disability.

Out of the need for dealing with the health problems of group living, there has evolved, with increasing clarity over the centuries, a recognition of the signal importance of community action in the promotion of health and the prevention and treatment of disease. This recognition and its consequences for action are summed up in the concept of public health. It has its roots in such rudimentary measures as the rituals and incantations against pestilence of ancient Mesopotamia and in the advanced administrative systems of imperial Rome, the ruthless isolation by the medieval community of the person afflicted with leprosy or plague, the needs of kings and princes to achieve and maintain power through large populations, as well as in the continuing efforts and devotion of numerous men and women who have desired only to better the condition of their fellows.

### Concepts and practices

Health and ill health have their biological roots, but the biological processes and phenomena have been and are being influenced, impeded, and facili-

tated in contexts of changing political, economic, social, and cultural elements. Indeed, the changing character of community health action is due to the circumstance that health problems do not exist in the abstract but have always been linked with the varying conditions of particular groups of people. One may almost say that each period of history has its own ideal of health and its own public health. For example, the concept of health in medieval Europe was limited by the theological concept of sin. This was particularly true in the case of mental health; not until the realm of health was extensively secularized was it even possible to begin to consider mental health in a public health context. In this sense, the concept of public health must of necessity be asymptotic. The goals of public health are not Platonic ideas, absolute and unchanging. They represent in some form the dominant values of a society, and public health activities have been explained or justified on several grounds, such as religion, humanitarianism, or scientific interest. An extremely pervasive consideration, furthermore, has been economic or social utility.

Historical analysis of public health reveals two major components that have been involved in its evolution. One is the development of medical science and technology. Understanding the nature and cause of disease provides a basis for prevention and control. Thus, faced with problems of epidemic disease, communities have acted in terms of some prevailing concept of the nature of disease. On the primitive level of knowledge, such action is generally couched in supernatural terms. For thousands of years epidemics were looked upon as divine judgments on the wickedness of mankind, and it was believed that these punishments were to be avoided by appeasing the wrathful gods. Alongside this theurgical theory of disease, however, there gradually developed the idea that pestilence is due to natural causes involving physical, biological, and social factors, that is, causes that can be studied rationally by the human mind. Beginning with the efforts of the Greeks, men have endeavored to combine speculations, theoretical inferences, observations, and experimental facts into theories that would explain the occurrence of diseases and provide a rationale for their prevention or control.

Effective action, however, depends on a variety of nonscientific elements, which are basically political, economic, and social. Disease may seriously disturb the life of the group. This is particularly true when people fall ill with contagious diseases and thus menace the health of their fellow men, or when individuals become a burden on the community because of illness or disability. To deal with such matters society creates agencies, establishes

laws, and institutes procedures to implement laws. Public administration, in a simple form, is found early in human history, and from the beginning public health has been closely linked with governmental activity.

Thus, the administration of public health is in large measure a political act. Any endeavor to understand public health practice must take account of the political dimension, because it lies at the base of the activity. Because public health is an expression of the aims and values of a society and is so closely linked with government, the practice of public health has been influenced not alone by the state of scientific knowledge but equally, if not more so, by prevailing theories of society and the state. The work of Johann Peter Frank in the eighteenth century is inconceivable apart from mercantilism and enlightened absolutism, just as that of Edwin Chadwick in the nineteenth century has its roots in philosophical radicalism and classical economics. Similarly, public health policy in the Soviet Union derives a characteristic stamp from communist doctrine, just as, in the United States, it has been molded by federalism, free enterprise, liberalism, and social reform.

### The Greco-Roman world

Evidence of activity related to community health has been found in the earliest civilizations. Bathrooms, drains, and water supply and sewerage systems testify to the achievements of the ancient Cretans, Egyptians, and others. Services today associated with public health are not mentioned frequently in antiquity; nevertheless, there were specific administrative arrangements for such matters as drainage and water supply, for which designated officials were responsible.

An important theoretical contribution of antiquity was the idea developed by the Greeks that disease is due to natural causes, involving the natural environment and especially climate. Ill health developed when there was an imbalance between man and his environment. This view is clearly expressed in the Hippocratic work *The Airs, Waters and Places*, the first known systematic endeavor to elucidate the causal relations between disease and environmental factors (climate, soil, water, mode of life, and nutrition). For more than two thousand years this basic approach provided the theoretical underpinning for an epidemiological understanding of endemic and epidemic disease. No fundamental change occurred in this respect until late in the nineteenth century, when the sciences of bacteriology and immunology were developed.

When Rome conquered the Mediterranean world

and took over the culture of the Greeks, it also accepted Greek ideas on health and disease but adapted them to Roman purposes. The Romans were no theoreticians, but they left their mark on history as engineers and administrators, builders of sewerage systems and baths, and providers of water supplies and other health facilities. Thus the waterworks of Rome were supervised by a board which had at its disposal a permanent staff of workers. Similarly, the maintenance and cleansing of the sewerage system was under the supervision of specially designated officials, who had a staff of public slaves. Equally significant was the organization of medical care. By the second century A.D. there was a public medical service, and hospitals had been created. These were first provided for soldiers, slaves, and public functionaries, later for civilians. Eventually, under the influence of Christianity, motives of charity and benevolence led to the creation of hospitals and related facilities in many localities. The foundation of hospitals for the sick, the disabled, and the indigent in the medieval period derives from Roman institutions.

Public health in Rome can clearly be seen as a social subsystem of the community with an organizational structure, a set of functions in terms of defined objectives, functionaries to carry out the necessary activities, a rationale to explain what was done, and techniques and tools for these purposes. Furthermore, this system had relationships with other parts of the social organization of Rome; the growth of the administrative system can clearly be seen as a consequence of the expansion of the community.

### The Middle Ages

The disintegration of the Greco-Roman world led to a decline of urban culture and with it to a decay of public health organization and practice. However, the East Roman, or Byzantine, Empire was to a large extent able to carry on the tradition and culture of the classical world. Here the Greco-Roman legacy was preserved, and from this center it was first transmitted to the Arabs in the East and later to the peoples of Europe. In the West during the earlier medieval period (A.D. 500-1000), communal activities in the interest of health were undertaken under the aegis of the church and particularly the monastic orders; they undoubtedly provided models for the urban communities that began to develop in Europe about the tenth century.

Many of the public health problems of the medieval community were simply due to the circumstance that an increasing population had to be accommodated within a limited space and that for a long time most of the inhabitants maintained a



rural mode of life. The problems were in essence those already indicated for antiquity, but to deal with them the medieval municipalities had to create anew the institutions needed for a hygienic mode of life. While the medieval community did not have an organized public health system in the present-day sense, it did have administrative machinery for disease prevention, sanitary supervision, and, in general, protection of community health. The character of this machinery was closely related to the general administration of the medieval municipality. The city or town was run by a council that had charge of finances, organized its provisioning, ordered and supervised public works, and also dealt with health and welfare problems. Physicians were not involved in public health administration but were employed for specific duties, such as the provision of expert counsel in times of pestilence or in medico-legal matters, the diagnosis of leprosy and similar conditions, and the provision of medical care to the indigent, to prisoners, or to other public charges as required.

**The institution of quarantine.** Medieval man was far from passive when faced by the problem of epidemics. He did what he could to protect himself, but in a manner consistent with the prevailing climate of thought and belief. Thus, his protective ideas were based on an amalgam of medical and religious views. The need to control leprosy was recognized early and led to a mode of public health action that is still with us, namely, the isolation of persons with communicable diseases. When the community believes itself menaced by such individuals it feels justified, acting through its institutions, in subjecting such people to restraints and sanctions in order to protect itself; people suffering from certain communicable diseases have to be reported to the authorities, and in some cases the freedom of the individual may be severely circumscribed. Thus, following policies adopted at the third Lateran Council in 1179, lepers were expelled from the medieval community, deprived of their civil rights, and consigned to a legal and social death. At the same time, places of refuge (leper or lazar houses) were provided for these unfortunate people.

The preventive principle used to combat leprosy was amplified and carried further in dealing with that other great scourge of the Middle Ages, bubonic plague, popularly known as the Black Death. From this preventive endeavor grew a basic contribution to public health practice, namely, the institution of quarantine. Beginning at Venice in 1348, public officials in Italy, southern France, and the neighboring area created a system of sanitary control to combat plague and other infectious dis-

eases, with observation stations, isolation hospitals, and disinfection procedures. Based on more accurate knowledge and organized more rigorously, quarantine is still a part of contemporary public health practice. Other areas of public health to which the Middle Ages made significant contributions were health education and the development of the hospital. Even though medieval hospitals had little in common with the modern institutions, they provided one of the sources from which our hospitals evolved.

### The modern era

With minor modifications, the public health pattern created by the medieval urban community continued in use from the sixteenth to the nineteenth century. With the development of national states in Europe, central governments took action increasingly but sporadically; on the whole, public health problems were handled by the local community. During this period, however, basic scientific knowledge was being acquired on which the structure of modern public health would eventually be erected. The great scientific outburst of this period began to make possible the more precise recognition of diseases and a better understanding of their nature. At the same time, ideological form was given to the possibility and importance of applying scientific knowledge to the health needs of the community. Concurrently, a quantitative approach to health problems developed out of the political and economic needs of the modern state. To increase national power and wealth, a large population was considered necessary. It was in relation to this concern that political arithmetic, that is, the collection and analysis of quantitative data bearing on national life, developed. The founder of this approach was William Petty, a seventeenth-century physician and economist; but the first solid contribution to vital statistics was made by his friend John Graunt, whose classic book *Natural and Political Observations . . . Upon the Bills of Mortality* appeared in 1662. The application of statistical analysis to community health problems was to prove extraordinarily fruitful for the development of public health.

The consideration of health problems in connection with the aim of maintaining and augmenting a population that could be economically productive and provide fighting men logically implied a health policy for the entire dominion of a monarch or of a nation. This implication was recognized and began to be developed in England and on the Continent during the seventeenth and eighteenth centuries. While a number of thinkers and men of affairs endeavored to deal with public health on a

national scale, it was not until the nineteenth century, with the advent of the new industrial and urban civilization, that the problem of organizing the larger community to protect its health became a matter of national concern and led to concrete results. The earlier efforts reached a high point in the work of Johann Peter Frank, whose career spanned the late eighteenth and early nineteenth centuries, and who is best known as a pioneer in public health administration and social medicine.

During the eighteenth century the cultural and economic movements known as the Enlightenment and the industrial revolution provided the seedbeds in which the new ideas, tendencies, and methods that revolutionized public health in the nineteenth century germinated and developed. In Britain, France, the countries of central Europe, and the United States, similar problems were attacked in the name of reason, order, human welfare, economy, and community concern. Among these were alcoholism, infant mortality, epidemics, the care of the insane, the creation or improvement of hospitals and dispensaries, improvement of the physical environment in towns, and the health conditions of specific groups such as soldiers, sailors, scholars, prisoners, miners, metalworkers, and various kinds of artisans. This interest is linked to the development of the health survey. The method was applied to the investigation of regions, communities, institutions, or population groups. Thus, John Howard published his account of the *State of the Prisons* in 1777 and proposed means to ameliorate social and health conditions which he had found. When applied to regions or communities, these investigations were known as medical topographies. Comprising sanitary surveys, epidemiologic studies, and social investigations, they prepared the way for the more specialized surveys and analyses that were carried out during the nineteenth and twentieth centuries. Today the survey as a tool for studying community health problems is an important component of the public health armamentarium.

### Public health and sanitary reform

Modern public health developed out of the sanitary reform movement of the nineteenth century, which began in England, where the impact of the industrial revolution on health was first recognized as a matter of community concern requiring governmental action on a continuing basis. Nevertheless, wherever industrialism developed, whether in France, Germany, or the United States, the consequences were similar and called for similar remedies. The human cost of industrialization and urbanization in terms of ill health and premature death was great, and the sanitary reformers en-

deavored to reduce it by organizing the community to protect the health of its members. This aim was coupled with a recognition that disease for which the individual could not be held responsible was an important factor in the cost of public assistance and that it would be good economy to undertake community-wide measures for the prevention of disease.

**In England.** The industrial revolution found England without any effective system of local government or any national agency to deal with the health problems which it created or intensified. At the same time urban communities grew and became more congested, and more and more people became aware of the cities' novel, powerful, and alarming qualities. It was apparent that endemic or epidemic diseases tended to seek out the poorer districts, but they were not limited to them. This awareness and the consequent desire to reduce or eliminate disease and the destitution which it produced are among the major roots from which the sanitary reform movement sprang.

One of the first products of this reform movement was the *Report on the Sanitary Condition of the Labouring Population of Gt. Britain*, a fundamental document dealing with modern public health which appeared in 1842. Prepared by Edwin Chadwick, lawyer, administrator, and one of the creators of sanitary reforms, the *Report* showed that communicable diseases were associated with filthy environmental conditions. In Chadwick's view, what was needed was an administrative organ to undertake a preventive program by applying engineering knowledge and techniques in an efficient and consistent manner. He also recognized the need for a physician to see that action was taken when necessary and to keep watch on health conditions in the community.

The Public Health Act of 1848, passed by Parliament after six years of agitation, established the General Board of Health. This step was a major landmark in the history of public health, because it created a basis for the further evolution of public health administration in England and led eventually to the establishment of the Ministry of Health in 1919. Moreover, it set an example whose influence was felt far beyond England. The impact of British sanitary reform was nowhere more pervasive than in the United States, where, as in England, one of the basic problems involved in the genesis and development of public health was the need to create an effective administrative mechanism for the supervision and regulation of community health.

**In America.** Between 1800 and 1830, only five major American cities had boards of health. Even



as late as 1875 many large urban communities had no health departments. Beginning in the 1830s, however, increasing immigration and urban growth produced situations that required urgent attention. Inadequate provision for housing, water supply, sewerage, and drainage brought into being a whole brood of evils that expressed themselves in the urban slum. Recurrent epidemics of smallpox, typhoid and typhus, cholera, and yellow fever impressed upon the public the urgent need for effective public health organization.

In 1866 the New York Metropolitan Board of Health was established, an event which marked a turning point in the development of American public health. This example was soon followed by other states and municipalities in establishing effective health departments: Massachusetts, 1869; California, 1870; District of Columbia, 1871; Minnesota, 1872; Virginia, 1872; Michigan, 1873; Maryland, 1874; Alabama, 1875; Wisconsin, 1876; and Illinois, 1877.

As state and local health departments were organized, the idea of a national health agency seemed the logical next step. However, it should be remembered that until practically the end of the nineteenth century the U.S. government had no concern with public health matters. Organization and action for the protection of community health were considered a local responsibility to be carried out by the state or the locality. Thus, the doctrine of state sovereignty continued to hold sway in the health field and handicapped public health action on a national basis for many years. Congress created the National Board of Health in 1879, but it was of little significance and disappeared after 1883. The establishment of a national health agency was not achieved in the United States until 1953, when the Federal Security Administration became the cabinet Department of Health, Education and Welfare, which included the Public Health Service, the Children's Bureau, and the Food and Drug Administration, as well as other health, welfare, and educational services of the federal government.

### Development of the health sciences

This evolution has gone hand in hand with the development of a complicated urban industrial society and has been one of the responses to the need for a more organized, efficient administration of health services. At the same time, the provision of a stable administrative structure has made it easier to incorporate new scientific knowledge into public health practice and to deal with new problems as they occur.

Public health action depends not only on govern-

mental organization and public attitudes but also on the health sciences and their technology. Understanding the nature and cause of disease provides a basis for preventive action and control. This was accomplished for many of the communicable diseases through the development of bacteriology and immunology, the sciences that have exercised the most profound influence on community health from the end of the nineteenth century to the present. It is almost impossible to overemphasize the consequences of this development. Based on the researches of Louis Pasteur and Robert Koch, as well as on the studies of their co-workers and successors, scientists identified the microorganisms responsible for specific diseases and uncovered their mode of action. The way was thus opened for public health action on a more specific, accurate, and rational basis. An example is the prevention and control of diphtheria through the use of antitoxin and toxoid.

Beginning about 1870 there was a continuing downward trend in mortality because of a decline in the frequency of certain diseases: chiefly smallpox, typhoid and typhus, yellow fever, malaria, and tuberculosis. These trends were roughly the same in the municipalities of western Europe and the United States and undoubtedly reflected first the impact of the earlier sanitary reform movement and then that of the bacteriological era. This period also saw an increasing concern with the health of mothers and children, problems of nutrition, health education, occupational health, and the organization and provision of health services on a community basis.

### The twentieth century—international trends

The past fifty years have witnessed an unprecedented over-all trend toward the improvement of community health. Yet, this advance has not been uniform either within communities or between various parts of the world. A large group of countries, generally underdeveloped in an economic and technologic sense, and often newly independent, still have problems of preventable disease like those with which the countries of western Europe and the United States had to cope one hundred years ago. Their problems are still the control of infectious diseases, the provision of uncontaminated water supplies and proper sewerage, and the elevation of the standard of living to a minimum acceptable level. In short, the underdeveloped areas of the world confront the twentieth-century public health workers with the same kind of problems that the sanitary reformers faced on a national scale in the nineteenth century.

International health action has grown out of a

broadening realization that in a world which for more than a hundred years has been contracting because of technological development and increasingly complex economic and political interdependence, the presence of disease in one area constitutes a continuing danger for many others. To deal with such problems, the World Health Organization was created in 1946, on the basis of experience obtained from earlier governmental and private organizations.

In the economically more advanced countries such as the United States and Great Britain, the actual problems of community health are different. As the diseases of infancy, youth, and early adulthood have been reduced so that people no longer die of them in great numbers, many more individuals live into the older years. As a result, among the current problems are the prevention or control of chronic conditions such as cancer, diabetes mellitus, arthritis, musculoskeletal diseases, and the mental changes associated with aging.

As problems of communicable disease have declined in urgency, community health programs have been altered to include other elements and conditions that may adversely affect the physical, social, and psychological well-being of people. In recent years the widening horizons of public health have brought to our attention problems of chronic disease, but also such problems as accident prevention, mental health, addictive diseases, the organization of medical care, and increasing needs for social services. Also important in the light of current problems is the renewed emphasis on control of the physical environment. Our expanding and changing technology has led to environmental alterations of increasing complexity. The once dominant problems of bacterially contaminated water and food have been replaced to a considerable degree by chemical pollution. In this category is the new and important field of radiation control.

The welfare state. These developments must be seen in the perspective of a world-wide historical evolution that has brought into being the modern state with its concern for individual, family, and community needs in health and welfare. There is probably no more fascinating process in recent history than that through which the *laissez-faire* "night watchman" state of the nineteenth century has been transformed into the present welfare state. The same broad developments occurred in all the leading industrial countries, although with numerous variations and, above all, differences in tempo due to varying historical traditions and conditions. Today, the principle of state intervention and control in health matters is accepted; the only difference is the greater or lesser efficiency of the

intervention and the means by which it is accomplished. Its emergence has resulted from the effects of important political, economic, and social trends. Thus, during this period, the typical trend of political and economic organization has been the continuous and progressive replacement of smaller units by larger ones. Another factor has been the need for national efficiency and the planned utilization of resources. These developments have necessarily led to the widespread acceptance of the need for a strong central authority entrusted with large powers to promote social well-being. This change has brought about an increasing rationalization and bureaucratization of health services, and it explains in part why in our time the social sciences have become an increasingly important element in public health.

The modern concept that a national government is responsible for the health of the people is an extension of the earlier view whereby the local community provided for such needs. As the center of power has moved from the small political unit to the large one, this shift has affected the provision of health services. Although the function of health promotion and protection is today lodged basically in the executive organ of the national community, the individual localities, as well as groups and persons in them, must still take an important part in the preservation of individual and collective health. Thus voluntary health agencies, that is, organizations not supported by tax funds, play an important role on the American health scene.

The relations of the national health service to local health organizations and personnel show wide variations throughout the world. In some countries, such as Great Britain and the Soviet Union, all health services are essentially socialized. In other countries, like the United States, the national health authority deals with international and interstate problems, carries out and stimulates extensive programs of research, encourages state and local health departments through financial support, and provides guidance where needed. Increasingly over the past forty years, the public health worker, in order to deal with the complex problems of contemporary communities, has had to become a highly trained, specialized professional. To provide the required education and training, institutions known as schools of public health have been established.

Public health has moved a long way from its beginnings. More and more, man can consciously plan for better health because available knowledge and resources make it possible in many instances to act with a clear understanding of what he is



doing. Many health problems have been solved basically, but knowledge awaits application in practice. In all countries there are problems of community health that require political and social action guided by available knowledge. In this sense, the dynamic and changing character of community health action and the significant trends and issues involved in it must be viewed as an aspect of the process of social change in society.

GEORGE ROSEN

[See also HEALTH; ILLNESS; MEDICAL CARE; MEDICAL PERSONNEL; VITAL STATISTICS; and the biographies of GRAUNT; PETTY.]

#### BIBLIOGRAPHY

- FISCHER, ALFONS 1933 *Geschichte des deutschen Gesundheitswesens*. 2 vols. Berlin: Rothacker.
- FRAZER, WILLIAM M. 1950 *A History of English Public Health: 1834-1939*. London: Baillière.
- GRAUNT, JOHN (1662) 1939 *Natural and Political Observations Made Upon the Bills of Mortality*. Baltimore: Johns Hopkins Press.
- GREAT BRITAIN, POOR LAW COMMISSIONERS (1842) 1965 *Report on the Sanitary Condition of the Labouring Population of Gt. Britain*, by Edwin Chadwick. Edinburgh Univ. Press.
- HANLON, JOHN J.; ROGERS, FRED B.; and ROSEN, GEORGE 1960 *A Bookshelf on the History and Philosophy of Public Health*. *American Journal of Public Health* 50:445-458.
- HOWARD, JOHN (1777) 1929 *The State of the Prisons*. New York: Dutton. → First published as *The State of the Prisons in England and Wales*.
- LESKY, ERNA 1959 *Österreichisches Gesundheitswesen im Zeitalter des aufgeklärten Absolutismus*. *Archiv für österreichische Geschichte* 122, part 1.
- PARISOT, JACQUES 1933 *Le développement de l'hygiène en France*. Nancy (France): Thomas.
- ROSEN, GEORGE 1958 *A History of Public Health*. New York: MD Publications.
- SIEGERIST, HENRY E. 1956 *Landmarks in the History of Hygiene*. London School of Hygiene and Tropical Medicine.
- SIMON, JOHN (1890) 1897 *English Sanitary Institutions*. 2d ed. London: Smith.
- SMILLIE, WILSON G. 1955 *Public Health: Its Promise for the Future*. New York: Macmillan.
- WINSLOW, CHARLES E. A. 1943 *The Conquest of Epidemic Disease: A Chapter in the History of Ideas*. Princeton Univ. Press.

#### PUBLIC INTEREST

The public interest is a concept that can be defined in several ways. It is often treated as an ideal standard to which appeals for validation of political claims and policies are directed. References to the public interest commonly occur in editorials, political speeches, and other utterances of an admon-

itory or edifying character. The term is also often encountered in judicial and regulatory pronouncements. The significance of the conception is normative rather than analytical.

#### Attempted definitions

In efforts made to define the term systematically, several categories appear. There is the formal meaning wherein the public interest is viewed as the objective of the duly authorized organs of government or as the expression of majority rule. In substantive or policy terms, the public interest may be envisaged as embracing those activities necessary to the safety of the state and the welfare of the community: defense, police protection, education, and public health and sanitation. The public interest has no a priori content waiting to be revealed. Public responsibilities regarded as being in the public interest in one setting may not be so regarded in a different context. The fact that the concept is elastic and relative rather than fixed and absolute makes it of greater utility in the quest for a supporting consensus as social changes occur and efforts at accommodation are made. The public interest then serves to remind the parties immediately concerned that there are considerations extending beyond their own goals or their particular rivalries or negotiations. Thus in labor-management disputes, or when subsidies or favorable tariff rates are sought for special groups, the consumer interest may be identified as the public interest. The individual may find himself in one role seeking to advance his self-interest and in another capacity allied with the larger good. Hence the term, whether for manipulative ends or for hortatory or inspirational purposes, is neither the device nor exclusive goal of any one group or class. As an aim to be articulated or an ideal to be enunciated, the public interest stands for the broad versus the narrow, the more inclusive versus the limited.

Indeed the very existence of the state as a collectivity would seem to prompt the formulation of basic common purposes. Thus the public interest is akin to the statements of ultimate ends that characterize political philosophy, varying from Plato's concern with the moral development of citizens as the purpose of the city-state to Hobbes's emphasis on order, Locke's protection of natural rights, and the utilitarian's assertion of the greatest good for the greatest number. As these illustrations indicate, the existence of a common interest is an implicit assumption, although its more specific manifestation cannot always be readily agreed upon. Yet both in traditional political thought and in contemporary organization theory, it has been noted

that despite uncertainty about exactly what the shared interests may be, some measure of sharing exists for every human association.

The public interest can be visualized as a continuum that represents the values, aspirations, and objectives of the community or polity. There are values that are clearly central, such as health; and there are objectives that are controversial, such as fluoridation of water or discouragement of cigarette smoking. There are aspirations that are nearly universal, such as world peace; but there are numerous policy objectives to this end, which are subjects for debate. It is impossible to state with precision where any given item belongs on the continuum, yet there is general acceptance that a range of activities and goals exists above irrational or selfish individual interests.

Granted that a wide consensus supports freedom, justice, and compassion, the problem is to objectify these ideals in public policies bearing concretely on such areas as civil rights, race relations, juvenile delinquency, or narcotics addiction. The public interest cannot be thought of as a compromise struck between the desires of dope addicts and moralists or between hoodlums and the police. The public interest is clearly more than the sum of competing interests. It is an insistent reminder that morality must have a central place in any society that respects the freedom of its members. The concept of public interest holds the ultimate ethical justification for the demands that the state makes on the individual. The high value placed upon freedom and justice calls for the readiness of the citizen to make sacrifices if necessary and to share responsibility for sustaining the values of the polity.

The concerns of the philosopher and the needs of the official and the politician will call for continuing reference to the public interest if for no other purpose than as a symbol, as a shorthand device for directing attention to considerations that transcend the immediate, the selfish, and the merely expedient.

The public interest, as a concept, seems to be employed in situations that involve a conflict of interests or a problem of defending or interpreting broader or longer-range considerations against special or more immediate factors. Thus T. H. Green in his *Lectures on the Principles of Political Obligation* refers to the "common good" in analyzing the rights and duties of the state and the citizen. Brief reference is made to the public interest in the following context: "If the injured individual were likely to fail in the institution of proceedings for his own redress or defence, the public inter-

est would require that the matter should be taken out of his hands" ([1882] 1960, p. 240; italics added).

Similarly Robert MacIver in *The Web of Government* refers to the general welfare and community, but when he describes the conflicting interests of modern industrial states, he writes that "government cannot afford to sit by while the disputants bring economic pressure to bear on one another. . . . In the public interest it must devise whatever measures are expedient . . . to prevent any interruption of vital services" ([1947] 1963, p. 350).

### Uses of the concept

The fact that the term is imprecise does not mean it is disappearing from ordinary usage; this is evident from the frequent references to the public interest in political polemics and in day-to-day discussions of public policy. The term then is justified as a useful working hypothesis: if it were abolished today, some other polar term would be coined to contrast the selfish goals of the few with the larger good, or sheer majority power with the proper rights of a minority. The occasional debasement and abuse of the term, it is argued, is no reason for its rejection (Colm 1960, p. 127). And as for ambiguity: Is "individual freedom" any more precise than "public interest"?

The concept of the public interest has won acceptance in the context of a functionally differentiated political system and a modern, industrial, and bureaucratized society that has a strong tradition of constitutionalism and rule of law.

The term is closely associated with the growth and professionalization of the civil service. The existence of a "neutral bureaucracy," officeholders neither subservient to the will of a monarch nor dependent upon a political party, calls for some focus of loyalty. The public interest helps to meet such a need.

Felix Frankfurter, in his *The Public and Its Government*, uses the concept in referring to the public officials "pitted" against the utilities experts, for whom they are no match "except for occasional men of great capacity and exceptional devotion to the public interest" (1930, p. 115). There are also occasions when officials are accused of exceeding their authority to the detriment of the public interest.

Where clearly distinguishable interests compete and policies can be seen as expressions of group desires, the public interest can be determined with some degree of specificity. Industrialization and improved means of communication and transportation have brought about conditions of complexity



requiring more competence and expertise to determine the public interest in particular cases.

In negative terms one can list the efforts made to protect the public interest from threats to public order, safety, and morals (laws and ordinances with respect to crime, accidents, vice) and in the economic sphere regulations applying to fraud, monopolies, vote fixing, fair employment practices, and consumer protection.

Institutionally and procedurally, legislative bodies must rely upon special agencies to discover and protect the public interest within their respective jurisdictions.

From the standpoint of jurisprudence it would appear that the concept has a firm position. According to one spokesman, "the lawmen have a vested interest in 'the public interest' as an operational concept: They would be tongue-tied without it" (Cohen 1962, p. 160). Nor have judges refrained from definition: "Public interest means something in which the public, the community at large, has some pecuniary interest, or some interest by which their legal rights or liabilities are affected" (*State v. Crockett*, 206 Pac 816, 817, quoted in Montgomery 1962, p. 222). And *Munn v. Illinois* states, "Property does become clothed with a public interest when used in a manner to make it of public consequence, and affect the community at large. . . . When . . . one devotes his property to a use in which the public has an interest, he . . . must submit to be controlled by the public for the common good" (94 U.S. 113, 126). Technological development and economic growth have resulted in industries whose activities go beyond the owners, shareholders, employees, competitors, and direct customers and affect the interests of a wide community, as in the case of the railroads and power companies.

### Recent scholarly discussion

The literature of recent years indicates a wide range of views concerning the significance of the concept. The differences of opinion result from the varying emphases with respect to its normative aspects and questions concerning its usefulness for analytical purposes. Since there is no generally accepted definition of the term, several different formulations are offered.

According to Walter Lippmann, "the public interest may be presumed to be what men would choose if they saw clearly, thought rationally, acted disinterestedly and benevolently" (1955, p. 42).

Alfred Verdross regards the public interest as

"the establishment of social conditions under which individual persons are able to build, through industry and productive work, a life which is in consonance with the dignity of the human being" (Bodenheimer 1962, p. 213).

Charles Frankel writes:

The ideal of the public interest calls on men, despite their egoism, to set their preferences side by side with the preferences of others and to examine them all with the same disinterestedness and impartiality. It asks them to seek as tolerable and comprehensive a compromise among these interests as is possible. And it reminds them that every decision they make is a limited one, that some interests may have been overlooked, that something better may be possible. (1962, p. 200)

Stephen Bailey (1962) calls public interest "the central concept of a civilized polity. Its genius lies not in its clarity but in its perverse and persistent moral intrusion upon the internal and external discourse of rulers and ruled alike" (p. 106). He also notes: "There is perhaps no better example in all language of the utility of myth than the phrase 'the public interest'" (p. 97).

The work of nearly a score of writers on the public interest is collected in *Nomos* No. 5 under the editorship of Carl J. Friedrich, who notes that it is not possible to discern "any distinctly philosophical, legal or political-scientific approach, though some of the authors lean more in one direction, some more in the other" (Friedrich 1962, p. viii).

Frank J. Sorauf (1962) has given thoughtful attention to "the conceptual muddle" concerning the "public interest"; his conclusions are as follows:

No single criterion or standard, such as the public interest, can subsume or encompass all that is good and desirable for society—regardless of the system of values one prefers or propounds. The matter of the goals of a contemporary industrial society consists in reality of a series of specific, policy-centered questions which involve the entire range of our concern, from a realistic approach to the emerging nationalism of Africa, through care for the ill among the aged, to the resolution of domestic racial conflict and tension. The answers to these policy questions will be hammered out painfully and pragmatically, and they will certainly be expressed in alternatives and wisdom considerably more specific and vastly less grandiose than the all-encompassing guise of a "public interest." If one argues that a public interest exists in all these problem areas, and if he can go the next step and identify it, all well and good. But why then resort to a concept of the public interest? (p. 188)

That the "public interest" has meaning for some public officials and interest groups is an incontestable fact. That it may in these terms affect the shaping and

administration of public policy is equally incontestable. Observation of American politics will also afford instances of its effective use as a unifying symbol and a social myth. Even should scholars reject the public interest for their own analysis, they must observe and record its prevalence and influence in the political system. But it is only as political datum that the public interest has a definable relevance to the study of politics and public policy. (p. 190)

Glendon A. Schubert has reviewed definitions and usage in an effort to discover whether the term could be used as a standard susceptible to empirical verification. His conclusion, if practicable, would point to a wide range of investigations:

If we assume that the peaceful adjustment of conflicting interests is not only the consummate art of the politician, but that it is also the fundamental task of all policy processes in a democratic polity, then a model of administrative due process would be empirically verified if, in practice, the decisions actually made resulted in the maximal accommodation of the affected interests, in comparison with the relative capacities of alternative structures for making the same decisions, and measured by a reciprocally minimal recourse to other centers for public policy change (i.e., the legislature, the chief executive, courts, etc.). . . . There would be technical problems in the construction of realistic models, and there would be practical problems in inducing political support for the translation into actuality of any models that went beyond the description of existing agencies and processes. (1957, p. 368)

Schubert concludes that in spite of these difficulties the task is worthwhile. In a later paper, however, he suggests that political scientists might do better to concentrate on concepts other than the public interest, concepts that promise more usefulness as scientific tools (1962, p. 176).

### Research needs and opportunities

From an analytical viewpoint, the concept of the public interest may be best understood when viewed within a specific context or continuum. There is ample opportunity for research into decisions of regulatory bodies and judgments by the courts where the concept is interpreted in the light of specific issues.

Inquiry into the purposes to which the concept is put, by whom, under what conditions, and with what consequences may lead to a clear understanding of the dilemmas and choice points that face opinion leaders, legislators, officials, judges, and all others who would engage in the contest of politics. A history of the usage of the concept might well be undertaken. What relationship can be

traced between the public interest and ideas of the general welfare or the national interest? Have significantly different shades of meaning or emphasis occurred? Moreover, international civil servants have yet to classify the *public* or *publics* they serve and the interests to which their energies are dedicated. Many of these problems can currently, be examined in the new "modernizing" countries where officials are uncertain of their status and their standards.

While a civil servant may feel that the public interest has been served if he follows the inner check of conscience or the course dictated by the highest technical or scientific standards of his profession, the problem of responsibility is not thereby resolved. The community's values are also involved in the public interest: public debate, discussion by the press, and legislative action or judicial intervention may be called for before an equilibrium is reached that appears to satisfy the articulate elements. There can obviously be no definitive adjustment "in the public interest" but rather a temporarily acceptable accommodation of interests and values. In this process of interaction there is opportunity for innovation and creativity and the public interest may be newly forged. The frontiers of scientific inquiry and problems of foreign aid in exotic cultures, for example, confront civil servants (and military officers) with necessities and opportunities for seeking fresh solutions for policy problems and unprecedented courses for governmental decision and activity.

A fresh perspective is gained if the public interest concept is viewed against the conditions that obtain in the developing countries. The problem of governance begins in the identification and articulation of interests. In traditional societies the peasant is a spectator or a pawn rather than a self-conscious participant. In the absence of effective interest groups, the charismatic leader rather than the politician qua broker has full opportunity. An underdeveloped society by definition lacks a well-defined infrastructure, and leaders accordingly are prone to make emotional appeals to a mass public and to stimulate nationalistic sentiments. Moreover, problems are not met through legal processes and administrative procedures of an essentially rationalistic character. Governmental sanctions are to be found in custom or in the will of the ruler or the authority of a dominant class rather than in appeals to the public interest. The ruler is the symbol or expression of power; governance is not thought of as a process of formulating issues and of analyzing, negotiating, manipulating, and adjust-



ing special interests to reach generally acceptable outcomes and thereby to articulate the public interest. As Lucian Pye writes:

Although in transitional societies there is generally a somewhat greater awareness of the potentialities of politics as a means of rationally solving social problems than there is in traditional systems, the expressive aspects of politics usually continue to occupy a central place in determining the character of political behavior. The peculiar Western assumption that issues of public policy are the most important aspect of politics, and practically the only legitimate concern of those with power, is not fully accepted in non-Western politics. Indeed, in most non-Western societies the general assumption is not that those with power are committed to searching out and solving problems, but rather that they are the fortunate participants in an exciting and emotionally satisfying drama. (1962, pp. 28-29)

Lippmann has argued (1955, p. 136) that "the art of governing well has to be learned. If it is to be learned, it has to be transmitted from the old to the young, and the habits and the ideas must be maintained as a seamless web of memory among the bearers of the tradition, generation after generation."

New nations, for example in Africa, need governmental institutions to provide justice and order. They need wise public policy to provide education, employment, and social welfare services for their rapidly growing populations. Here are newly enfranchised voters aspiring to the fruits of modernity but arriving at the status of citizenship without the historical experience or traditions out of which grew the polities of the West. Can the concept of the public interest be introduced; or can these new societies get along without such a unifying and disciplining concept? One answer offered is the theory of "one party rule." Loyalty to the leader of a single party can be quickly though not always permanently imposed, as Nkrumah and many other leaders have demonstrated. Acceptance of the public interest as a guiding belief would mean restraint on the ambitions of politicians and limits to the demands of the electorate.

In the developing countries, as industry advances, problems of administrative discretion can emerge, but whether the public interest will become an influential concept is uncertain. Will the civil service be motivated to serve the public? Will the courts be called upon to bring large economic enterprises within their jurisdiction? Will a recognition of individual rights requiring adjudication between adversary parties lead judges to apply criteria of public interest? Will the ties of kinship or loyalty to village tribe, community, or region permit an

awareness of the larger configuration of "public"? Will the acceptance of tradition stand in the way of the rational pursuit of "interest"? Does the concept strengthen nationalism? Dictatorship?

The answers to such questions will determine whether the term *public interest* has a universal applicability or whether it will remain a feature of Western jurisprudence and political thought.

The rule of law, due process, a free press, a loyal opposition, and the public interest are all value-laden concepts the limits and substance of which are difficult to define with precision, but all are significant in the maintenance of democratic government. Sport without the idea of fair play would lose zest—so, too, politics bereft of appeal to the public interest. The task of the umpire, in both situations, is probably made easier psychologically by reliance on such abstractions; but the responsibility of decision is in no way lessened, and the readiness of participants to question and of other observers to object remains.

The public interest is a concept that leaves open the way to change, as general acceptance for new policies is won and officials remain accountable for their decisions. Since there are many ways for seeking public acceptance and for safeguarding administrative responsibility within a democratic polity, it is doubtful that anything would be gained by insistence upon a precise definition of the public interest.

PENDLETON HERRING

[See also BUREAUCRACY; CIVIL SERVICE; CONFLICT OF INTEREST; DEMOCRACY; GENERAL WILL; NATIONAL INTEREST; PUBLIC POLICY; RESPONSIBILITY.]

#### BIBLIOGRAPHY

- BAILEY, STEPHEN K. 1962 The Public Interest: Some Operational Dilemmas. Pages 96-106 in Carl J. Friedrich (editor), *The Public Interest*. Nomos No. 5. New York: Atherton.
- BODENHEIMER, EDGAR 1962 Prolegomena to a Theory of the Public Interest. Pages 205-217 in Carl J. Friedrich (editor), *The Public Interest*. Nomos No. 5. New York: Atherton.
- COHEN, JULIUS 1962 A Lawman's View of the Public Interest. Pages 155-161 in Carl J. Friedrich (editor), *The Public Interest*. Nomos No. 5. New York: Atherton.
- COLM, GERHARD (1960) 1962 The Public Interest: Essential Key to Public Policy. Pages 115-128 in Carl J. Friedrich (editor), *The Public Interest*. Nomos No. 5. New York: Atherton.
- DOWNS, ANTHONY 1962 Public Interest: Its Meaning in a Democracy. *Social Research* 29:1-36.
- FRANKEL, CHARLES 1962 *The Democratic Prospect*. New York: Harper.
- FRANKFURTER, FELIX (1930) 1964 *The Public and Its Government*. Boston: Beacon.
- FRIEDRICH, CARL J. (editor) 1962 *The Public Interest*. Nomos No. 5. New York: Atherton.

- GREEN, THOMAS HILL (1882) 1960 *Lectures on the Principles of Political Obligation*. London: Longmans. → Published posthumously.
- HERRING, PENDLETON 1936 *Public Administration and the Public Interest*. New York: McGraw-Hill.
- LEYS, WAYNE A. R.; and PERRY, CHARNER M. 1959 *Philosophy and the Public Interest*. Chicago: Committee to Advance Original Work in Philosophy.
- LIPPMANN, WALTER 1955 *Essays in the Public Philosophy*. Boston: Little.
- MACIVER, ROBERT M. (1947) 1963 *The Web of Government*. New York: Macmillan.
- MEYERSON, MARTIN; and BANFIELD, EDWARD C. 1955 *Politics, Planning, and the Public Interest: The Case of Public Housing in Chicago*. Glencoe, Ill.: Free Press. → A paperback edition was published in 1965.
- MONTGOMERY, JOHN D. 1962 Public Interest in the Ideologies of National Development. Pages 218-226 in Carl J. Friedrich (editor), *The Public Interest*. Nomos No. 5. New York: Atherton.
- PYE, LUCIAN 1962 *Politics, Personality, and Nation Building: Burma's Search for Identity*. New Haven: Yale Univ. Press.
- SCHUBERT, GLENDON A. 1957 The Public Interest in Administrative Decision-making: Theorem, Theosophy, or Theory? *American Political Science Review* 51: 346-368.
- SCHUBERT, GLENDON A. 1961 *The Public Interest: A Critique of the Theory of a Political Concept*. New York: Free Press.
- SCHUBERT, GLENDON A. 1962 Is There a Public Interest Theory? Pages 162-176 in Carl J. Friedrich (editor), *The Public Interest*. Nomos No. 5. New York: Atherton.
- SORAUF, FRANK J. 1957 Public Interest Reconsidered. *Journal of Politics* 19:616-639.
- SORAUF, FRANK J. 1962 The Conceptual Muddle. Pages 183-190 in Carl J. Friedrich (editor), *The Public Interest*. Nomos No. 5. New York: Atherton.

## PUBLIC LAW

*The articles under this heading deal with public law as a field of study. A detailed guide to related articles can be found under LAW.*

### I. THE FIELD

### II. COMPARATIVE PUBLIC LAW

David J. Danelski  
Stefan A. Riesenfeld  
and Gerhard Casper

### I

### THE FIELD

Public law, viewed in its broadest terms as a field of inquiry in political science, embraces the study of all legal phenomena connected with politics or government. Whether specific phenomena—e.g., theories, rules, institutions, processes, and behaviors—are to be characterized as "legal" or "connected with politics or government" depends upon how law, politics, and government are defined.

An earlier conception of the field, which derived from the work of European legal scholars, was

somewhat narrower. The terms *public law*, *droit public*, and *öffentliches Recht* were understood as implying a functional dualism of the two main branches of positive law—private and public. Subdivisions of the former—e.g., commercial law—were viewed as dealing with the resolution of conflicts among individuals, and subdivisions of the latter—e.g., criminal, administrative, and constitutional law—were viewed as dealing with the resolution of conflicts between individuals and government, governmental regulation of individual and corporate activity, and the powers and limits of government (Walz 1934).

American public law scholars have had diverse conceptions of their field. Some have viewed public law as identical with constitutional law; others have seen it as limited to the study of courts and judges; and still others have taken the view that it covers all rule making and rule interpretation intended for the guidance of governmental agencies. Noting these competing conceptions of the field, Swisher (1946) reported that most specialists saw no advantage in formulating a theoretical definition that would isolate the field's distinguishing characteristics. They were content to define public law in terms of the areas in which they taught and did research, such as constitutional law, administrative law, jurisprudence, and judicial biography. Topics like civil liberties, judicial processes, and judicial behavior were later added on the same basis.

Closely related to practically every other field in political science, public law is one of the core fields of the discipline. Given the legal orientation of modern societies, there are few areas of government that do not have legal dimensions. Thus, constitutional law is linked with national and comparative politics, administrative law with public administration, jurisprudence with political theory, judicial behavior with political behavior, and so forth. When the American Political Science Association was established in 1903, the central position of public law was clearly recognized in the association's stated purpose—"the scientific study of Politics, Public Law, Administration and Diplomacy."

Public law is as old as political science itself, and at times the two have been almost indistinguishable. In the nineteenth century and early in the present century, the work of Europeans dominated the field. Krabbe, Duguit, and Laski are still familiar names in public law. But in the past several decades innovation and development in the field have occurred, for the most part, in the United States, where modern political science has made



greatest strides as an autonomous discipline. For that reason, the work of American political scientists since World War II is emphasized.

### Approaches

**The analytical approach.** The influence of *Staatslehre* thought lay heavily on public law in the closing decades of the nineteenth century. Law was viewed as the command of the sovereign, and sovereignty was ascribed to the state, source and enforcer of all law. Analytical scholars explored the "true nature" of the state and constructed elaborate theories of sovereignty. Analysis was formal and deductive; its main purpose was to give coherence and logical consistency to accepted legal doctrines.

The works of W. W. Willoughby (esp. 1896; 1924) are prime examples of the analytical approach. Influenced by the writings of Austin, Jellinek, and Laband, he devised a highly abstract, juristic theory of government. Ironically, because of the rise of pragmatism and realism in political science early in Willoughby's career, his work was already something of an anachronism during his lifetime. Nevertheless, it was taken seriously and discussed even as late as the 1940s (Cairns 1935; Cole 1948). But long before that, most political scientists had rejected Willoughby's abstract theoretical writings as sterile legalism, and today they are of little more than historical interest [see WILLOUGHBY]. As in the earlier period, when the analytical approach was dominant, the significant contributions in this recent area have come from European-trained legal scholars (e.g., Kelsen 1945; Hart 1961), not from political scientists [see KELSEN].

**The juridical approach.** Although it stems from analytical jurisprudence, the juridical approach is relatively unconcerned with general abstract theories of law, government, or the state. It views law primarily as a system of concepts, rules, and principles that are supposedly the basis of legal decisions. It focuses on constitutions, statutes, judicial decisions, and related phenomena, which are described, clarified, analyzed, and sometimes evaluated. Analysis emphasizes legal reasoning and probes the logic and consistency of legal decision making. Since the approach is concerned with the formal results of decision making, usually in terms of legal doctrine, it is necessarily formal; nevertheless, it can be tempered with considerable political realism.

The rise of the juridical approach and pragmatism coincide, and some juridical scholars (e.g., Powell 1918) clearly were influenced by the writ-

ings of pragmatists like John Dewey. Early in the twentieth century the foremost juridical scholar was Frank J. Goodnow, whose works in comparative and American administrative law (1893; 1905) are considered classics. His work was highly descriptive and based on extensive case analysis, yet it contained an incipient political realism, especially noticeable in his analysis of American constitutional law (1911). And in this regard, his teaching was as important as his writing. "Goodnow," wrote Haines and Dimock, "was constantly telling his students that government could not be understood and lasting improvements could not be made unless the political behavior and the political motivation underlying and eventually controlling constitutional and administrative law were properly mastered" (1935, p. xiii). [See GOODNOW.]

Goodnow's students—Thomas Reed Powell and Charles Grove Haines, among others—carried forward political realism in public law. Together with Edward S. Corwin, they called attention to the fact that the Supreme Court of the United States, in interpreting the constitution, was making public policy. Powell, who emerged as the leading juridical scholar of the 1920s and 1930s, brilliantly analyzed the logic and rhetoric of Supreme Court opinions and showed that at times personal preference, not ineluctable legal logic, was the basis of constitutional decisions. Haines (1922) went even further and suggested, in an article now considered a forerunner of judicial behavioralism, that constitutional law might be systematically explained in terms of psychological motivation.

It must be emphasized that not all juridical scholarship is characterized by political realism. When Powell and Haines were writing, many juridical scholars were doing formal legal analysis of judicial decisions, and, to a lesser extent, this is so today.

The juridical approach reached its peak during the interwar years. Public law scholars of the period concentrated on constitutional law. Their work was published in leading law reviews as well as in political science journals; they were sought as members of law faculties; and a few of them were given legal posts in government. Powell, who had been trained in law before taking his doctorate in political science, spent most of his career teaching at the Columbia and Harvard law schools. Walter F. Dodd, a political scientist without a law degree, joined the Yale law faculty in the 1920s. And Corwin and Carl B. Swisher, who also had no formal legal training, were special assistants to the U.S. Attorney General in the 1930s. Perhaps the most significant tribute to public law specialists

of the period was the prominence of their work in the *Selected Essays on Constitutional Law* (1938), which was edited by a committee of the Association of American Law Schools. About 40 per cent of the signed articles in the work were by political scientists—Powell and Corwin led all others with 23 and 14 articles, respectively. The work of Cushman, Haines, Dodd, Grant, Fairman, and Field also was represented. During this period juridical scholars were also highly esteemed in political science: Corwin, Powell, Haines, Cushman, and Dodd all became presidents of the American Political Science Association.

The juridical approach is still significant in public law. Constitutional law, usually taught by the case method, is the main public law course in American colleges and universities. As in the past, leading juridical scholars—e.g., David Fellman, William M. Beaney, and John P. Roche—still publish much of their work in law reviews. But the era of intellectual leadership by political scientists in constitutional law is over. Leadership has now passed to law professors, many of whom have been law clerks to Supreme Court justices.

The quality of juridical analysis has often been said to be the best in American political science. "If our scholarship in this general area is worthy of high praise," wrote Hyneman (1959, p. 44), "it is because of distinguished achievement in describing, evaluating, and criticizing public policies which interpret, extend, and modify the language of the United States Constitution."

Juridical scholarship has relevance beyond its contribution to the literature of political science. Corwin once said that if judges make law, so do legal commentators. Articles and books by juridical scholars are often cited in appellate briefs, arguments, and opinions. Thus their work in clarifying, criticizing, and analyzing legal doctrine is useful and perhaps influential in the judicial process. Mere citation of judicial scholarship in judicial opinions does not prove influence, but it does indicate that political scientists are often participants in, as well as observers of, the judicial process.

**The historical approach.** Early in his career, Corwin identified his approach to public law study as primarily historical. He took the position that because law is so avowedly historical in its manner and growth, it has to be understood historically. In practically every area of the historical approach, Corwin did pioneering work, and in many of these areas his contributions remain unsurpassed.

His conception of law showed the influence of historical jurisprudence, but it differed little from the juridical conception. This has been generally true of

historically oriented public law specialists, many of whom were students of Corwin's. The principal difference between the historical and juridical approaches lies not so much in the conception of law as in the method and the universe of phenomena regarded relevant for study. The historical scholar's forte is genetic explanation, which, at its best, is based on sound historical method—use of primary sources, critical evaluation of evidence, and conclusions drawn strictly from the evidence. If he is interested in the American constitution, he will, like the juridical scholar, study judicial opinions, but his universe of phenomena embraces much more—public opinion, group activity, presidential decisions, congressional behavior, and anything else that may account for constitutional growth (see Swisher 1943).

Historical scholarship in public law has included legal ideas (e.g., Corwin 1928–1929), constitutional development (e.g., Swisher 1943), constitutional concepts and doctrines (e.g., Wright 1938), judicial institutions (e.g., McCloskey 1960), and case studies (e.g., Westin 1958). But the area that has received greatest attention since 1930 has been judicial biography.

Judicial biography began in an important way in the United States with the publication of the first two volumes of Beveridge's *Life of John Marshall* (see 1916–1919), but Corwin's brief biography (1919) of the same judge was equally important, for it set the model most judicial biographers in political science were to follow. The typical judicial biography, as distinguished from studies of the legal "philosophy" of judges, is based on private papers of the subject and other primary sources. The judge's childhood is usually covered briefly; his legal and political career is covered in greater detail; and the bulk of the biography dwells on the judicial period, which gains focus from the judge's opinions, private correspondence, conference memoranda, diaries, and intracourt communications. All of this is usually set against and woven into the political history of the period. When it is done well, the biography contributes significantly to the understanding of an era of history. Even when it does not rise to such heights, it often provides insights into judicial behavior and decision making. Among the political scientists who have written significant judicial biographies are Swisher (1935), Fairman (1939), Mason (1956), and Magrath (1963). Journalists, lawyers, and law professors have also written judicial biographies similar in approach to those of political scientists, and it appears that leadership in this area, as in constitutional law, is passing to those law profes-



sors who have been law clerks of Supreme Court justices, for they often have special access to necessary private papers, as well as personal experience with the judges about whom they write.

Historical scholarship in public law has often received high praise from historians. Some political scientists, such as Swisher and Mason, have used historical method so well that they are acknowledged as leading American constitutional historians. Criticism of historical scholarship has come largely from political scientists who maintain that such work is atheoretical and yields no systematic generalizations about legal and political phenomena. Although this is a fair characterization of most historical work in public law, the criticism is inapplicable to the historical approach generally, for historical inquiry can be informed by theory and can yield systematic generalizations. Such inquiry in public law, however, has barely begun.

Historical scholarship can play an important role in the judicial process. That is what Holmes had in mind when he wrote that "a page of history is worth a volume of logic" (256 U.S. 345, 349), and that is why Cardozo (1921) designated history as a conscious method of judicial decision making. Interpretations of constitutional provisions often turn upon historical assumptions that research can clarify and sometimes show to be unfounded. With a few exceptions (e.g., Graham 1938), relatively little work has been done in this area.

**The ethical approach.** The ethical approach raises the question of justice in regard to legal rules, decisions, behaviors, and similar phenomena. It is invariably taken by those who have an ideal conception of law, e.g., the proponents of natural law, who maintain that an unjust law is no law at all. But the ethical approach can also be used by those having other conceptions of law.

For centuries the leading ethical approach has been natural law, which asserts the existence of an absolute, transcendent standard knowable through the use of reason [see NATURAL LAW]. Except for studies of natural law as a historical phenomenon (e.g., Corwin 1928-1929; Haines 1930), American public law scholars have had little to do with it. This was probably because of the field's early commitment to analytical jurisprudence. The only major study using an ethical approach that comes close to natural law is Berns (1957).

The question of justice has been dealt with in terms other than natural law [see JUSTICE]. Cardozo (1921; 1924; 1928) defined justice in terms of social values empirically ascertained. Brecht (1959) discussed justice similarly and indicated how ethi-

cal questions concerning law and politics might be approached in terms of scientific method. Brecht's work opens a fruitful area of public law scholarship yet to be developed [see BRECHT].

**The group-process approach.** Articulated initially by Arthur F. Bentley (1908), the conception of law underlying the group-process approach is that of a multitude of complexly related activities indicating the formation, systematization, struggle, and adaptation of group interests. Law, according to Bentley, does not result from government; it is the same phenomenon, "only stated from a different angle" ([1908] 1949, p. 272). What distinguishes legal and governmental activity from all other activity is its connection with governmental agencies. "The law at bottom," wrote Bentley, "can only be what the mass of the people actually does and tends to some extent to make other people do by means of governmental agencies" (*ibid.*, p. 276). Courts are obvious governmental agencies, but their activity, Bentley maintained, can be explained only in relation to all other connected activity—i.e., the group struggle—and not in terms of logic, ideas, or theories propounded by judges in their official opinions.

The implications of Bentley's conception of law were enormous, for it expanded the universe of phenomena relevant to public law to include the activity of criminals, policemen, prosecutors, lawyers, trial and appellate judges, law clerks, and anyone else whose actions are connected with the legal-governmental process. For its time, Bentley's work was an important contribution. It was a kind of sociological jurisprudence, but it differed from Pound's ideas in that area and was more advanced than Ehrlich's work. Thus, it is almost incredible that Bentley's work was ignored for more than two generations.

Bentley's eventual "rediscovery" was by a law professor, Karl N. Llewellyn. "Bentley," Llewellyn wrote, "saw and said in 1908 all that should have been necessary to force constitutional theory into total reconstruction" (1934, p. 1). Thereafter some political scientists (e.g., Shepard 1939) also recognized Bentley's contribution to public law, but another generation went by before it had a substantial impact on the field. Then, in the early 1950s, a number of group-process studies touching public law were published (Truman 1951, chapter 15; Latham 1952; Peltason 1953; 1955). [See BENTLEY; LLEWELLYN.]

Peltason's work had a special relevance to public law because he emphasized the role of courts and judges as participants in the group process. He

iterated and clarified Bentley's basic ideas and used them as the foundation for two studies concerning the federal judiciary (1955; 1961). Others, following his lead, extended group-process analysis to the state judiciary (e.g., Jacob 1965) and to the impact of judicial decisions on society (e.g., Patric 1957). The rise of the group-process approach in public law also stimulated a number of other studies, which, although not emphasizing Bentleyan thought, viewed interest groups or the group process as central (e.g., Vose 1958; 1959; Grossman 1965).

The group-process approach has grown in importance since the 1950s. Its main contribution has been the description of a broad universe of legal phenomena. But in theoretical terms, group-process studies in public law have gone little beyond Bentley's basic insights in 1908. The future importance of the approach turns on whether it is able to offer sophisticated explanation based on verifiable theory, and not merely careful description of legal-governmental processes. [See POLITICAL GROUP ANALYSIS.]

The behavioral approach. The behavioral approach seeks to explain legal phenomena in terms of operationally defined concepts and empirically verifiable hypotheses concerning human behavior. Focusing on the behavior of individuals in legal processes, it draws heavily on individual and social psychology. The approach involves prediction, usually understood in terms of probability; i.e., if phenomena X, Y, and Z are observed in some postulated relationship, then the probability is that R', not R'', will occur (see Lasswell & McDougal 1943, p. 238). In order to confirm hypotheses underlying such a statement and to determine probability, quantitative methods are often used, but quantification per se does not characterize the behavioral approach.

Political and legal realism (e.g., Haines 1922; Frank 1930) were forerunners of behavioralism in public law, but the approach received greatest impetus in its early years from Charles E. Merriam. Critical of narrow legalism, he urged political scientists to use psychology, psychopathology, sociology, and statistics in their research. Statistical study of the judiciary, he pointed out in 1921, was a research frontier that could be profitably penetrated. He exhorted public law specialists to spend less time in the law library reading cases and more time doing empirical research (1926). And at the University of Chicago, Merriam provided the conditions for creative behavioral work [see MERRIAM]. Mott, one of his colleagues, responded early to his

call for empirical research in public law (see, e.g., Mott et al. 1933). But Merriam's greatest impact in public law came in the 1940s, through the work of two of his former students and colleagues, Harold D. Lasswell and C. Herman Pritchett.

Lasswell's important work in public law began soon after he joined the Yale law faculty. In an article with McDougal (1943), he presented a profound criticism of the juridical approach (which he characterized as "syntactics"), sketched the outlines of a theory of judicial decision making, and indicated how the study of legal phenomena might be approached as a policy science. Although the article had an impact in legal circles, political scientists in public law manifested greater interest in one of Lasswell's later works (1948), in which he presented psychoanalytical case histories of three judges (Judges X, Y, and Z) to illustrate basic political character types. Lasswell's work in public law perhaps has not yet been fully appreciated. It appears that his influence in this area has been greater in law than in political science.

Modern behavioralism in public law began with the first of Pritchett's pioneering studies (1941) that identified and described bloc voting in the Supreme Court of the United States by means of quantitative analysis of nonunanimous decisions. His novel methodology stemmed from the work of another Merriam student, Herman C. Beyle (1931), but Pritchett's contribution was not just methodological. His study of the Roosevelt court (1948) skillfully blended behavioral, juridical, and historical approaches. Although it was one of the most innovative studies in more than a generation, initial response to it was critical on the ground that quantitative analysis was inappropriate in the study of the Supreme Court. Pritchett's contribution then lay fallow for almost a decade, and even he began to move away from quantitative analysis (see 1954). But by the mid-1950s a number of scholars (e.g., Bernard 1955; Kort 1957; Schubert 1958) had become interested in the quantitative study of the judiciary, and judicial behavior became firmly established as a subfield of public law.

After Pritchett, the study of judicial behavior owes more to Glendon Schubert than to any other scholar. Influenced by the work of Pritchett, and also by Beyle, Schubert developed a number of quantitative methods others had explored, e.g., bloc analysis, cumulative scaling, and factor analysis. He was the first to use game theory in the analysis of judicial behavior and was also the first to set forth a psychometric model of judicial decision making. A prolific scholar, he wrote or edited most



of the basic books in judicial behavior in the late 1950s and early 1960s (see, e.g., 1965). His review of the subfield's literature (1963) comprehensively covers the work of such other judicial behavioralists as S. Sidney Ulmer, Joseph Tanenhaus, Harold Spaeth, John R. Schmidhauser, Stuart Nagel, Fred Kort, and David J. Danelski. [See JUDICIARY, *article on JUDICIAL BEHAVIOR*.]

Judicial behavioralism has been criticized on the grounds that its methods have been too crude to measure the subtle art of judging (Mendelson 1963), that it has ignored the question of justice (Berns 1963), and that often its use of social science theory and methods has been improper (Becker 1965). It is true that some of the early behavioral work seemed crude because it was empiricism uninformed by theory, and when theory developed elsewhere was employed, it was sometimes used rather loosely. Nevertheless, there is nothing peculiar about judicial decision making that makes it less susceptible to behavioral analysis than are other kinds of subtle political decision making. And such analysis can even be used, at least in a limited way, in dealing with questions of justice. As the behavioral approach becomes theoretically sophisticated and is used in a number of areas of public law, current criticisms are apt to be obviated.

The work of behavioralists in public law is potentially useful to participants in legal processes. If, for instance, judicial decision-making theory develops to the point where it has high predictability, it will probably be used by lawyers in deciding whether to settle or appeal individual cases. Such theory, of course, will also be useful to judges in understanding their own behavior.

### Evaluation

**The field in perspective.** Although public law has responded to the major intellectual currents in law and political science, its response has often been late and then only partial. The field has given up old approaches reluctantly and has sought to accommodate the old with the new rather than to attempt bold reconstruction. That is illustrated by the juridical approach's growing out of its analytical predecessor and making only mild concessions to pragmatism and realism. One of the reasons for the criticisms of the behavioral approach is that it is not susceptible to easy integration in the field. To accept behavioralism entirely would mean that much, though not all, in the older approaches would be forsaken. Given the field's historically conservative attitude in such matters, most of its

specialists are unwilling to go that far. Indeed, it is probably because of its conservatism that public law has in recent years followed rather than led the discipline.

Public law in the United States has been preoccupied with the Supreme Court and its decisions. The court is prominent even in the work of those who use the group-process approach (see Vose 1959; Peltason 1961). That is not surprising, for the Supreme Court is a public law phenomenon par excellence; it is simultaneously a legal and a political institution, making decisions that have a corresponding dual character. Although studies of the Supreme Court and its decisions have been the core of public law in the United States, there has always been some interest in public law phenomena in the states and in foreign countries (see, e.g., Mitau 1965; Grant 1958). The approaches taken have usually been juridical or historical (e.g., Bayley 1962), but recently the behavioral approach has also touched this area, and scholars in a number of countries have begun to study judicial behavior much in the manner of American behavioralists (e.g., Hayakawa 1962).

An examination of the rise of various approaches indicates a trend away from formalism toward rigorous empiricism. This was due, apparently, to successive responses to pragmatism, political realism, legal realism, and behavioralism. The only major exception is the work of Berns (1957), which has not started a countertrend. On the contrary, the trend in ethical analysis in public law appears to be consistent with modern scientific method (see Brecht 1959).

Despite public law's importance in political science during certain periods of its history, it is difficult to point to significant theoretical contributions. To be sure, there was Willoughby's juristic theory and there were Bentley's insights concerning the legal-governmental process. But in regard to other areas of public law scholarship the theoretical contributions have been slight. Behavioralists are now developing microtheories of judicial decision making, but whether their work will be significant remains to be seen. All in all, the greatest single failing in public law appears to have been in the area of theory.

**The current state and prospects.** Public law has fallen from the high place it once occupied in political science. In terms of significance of its recent work, a survey (Somit & Tanenhaus 1964) has shown that it ranks lower than any other field in political science except political theory. Also, in recent years there has been a relative decline in

the number of political scientists entering the field. These phenomena are probably related to the field's failure to keep abreast with the discipline, deep divisions within it concerning approaches and assumptions, and the fact that today an adequate training in public law requires proficiency in practically all of its approaches.

Nevertheless, there are some indications that the field is experiencing a resurgence. In the early 1950s, a committee of the American Political Science Association expressed concern about the neglect of public law in the discipline, and not long thereafter one public law specialist was ready to concede that the field had "all but disappeared as a recognized object of study of political science" (Sherwood 1958, p. 87). Despite such estimates, the field is very much alive today. The group and behavioral approaches, if they did nothing else, at least brought public law into the main stream of political science.

There have been some attempts at integration of approaches. Some public law specialists have taken the position that for certain purposes all of the approaches can be useful and that much can be learned by integrating old and new approaches. For example, Johnson (1965) has used group-process and behavioral approaches in his study of the impact of Supreme Court decisions; Danelski (1961; 1964) and Murphy (1964) have used historical and behavioral approaches in their studies of the Supreme Court.

Public law is a field in transition. Convergence of approaches will facilitate transition, but the direction of movement seems clear. The field will acquire a measure of unity in a sophisticated, broad-gauged behavioralism that will take into account knowledge gained through the use of the other approaches. The change will be evolutionary, and traditional approaches, somewhat modified, will probably be used for some time to come.

As the field moves toward the behavioral approach, it is likely to make theoretical contributions that will have significance beyond the study of public law. If important theoretical contributions are forthcoming from the study of collegial judicial decision making, they will be useful in the study of legislative committees, administrative bodies, and other similar decision-making groups. The highest achievement of the field would be the development of a general theory of public law. If that is accomplished, public law will perhaps regain the high place it once occupied in political science.

DAVID J. DANELSKI

[See also ADJUDICATION; JUDICIAL PROCESS; JUDICIARY; JURISPRUDENCE; LAW; LEGAL REASONING; LEGAL SYSTEMS; POLITICAL SCIENCE.]

#### BIBLIOGRAPHY

- ASSOCIATION OF AMERICAN LAW SCHOOLS 1938 *Selected Essays on Constitutional Law*. 4 vols. Chicago: Foundation Press. → A supplement was published in 1963 by West as *Selected Essays on Constitutional Law, 1938-1962*.
- BAYLEY, DAVID H. 1962 *Preventive Detention in India: A Case Study in Democratic Social Control*. Calcutta: Mukhopadhyay.
- BECKER, THEODORE L. 1964 *Political Behaviorism and Modern Jurisprudence: A Working Theory and Study in Judicial Decision-making*. Chicago: Rand McNally.
- BENTLEY, ARTHUR F. (1908) 1949 *The Process of Government: A Study of Social Pressures*. Bloomington, Ind.: Principia Press.
- BERNARD, JESSIE 1955 *Dimensions and Axes of Supreme Court Decisions: A Study in the Sociology of Conflict*. *Social Forces* 34:19-27.
- BERNS, WALTER 1957 *Freedom, Virtue, and the First Amendment*. Baton Rouge: Louisiana State Univ. Press.
- BERNS, WALTER 1963 *Law and Behavioral Science*. *Law and Contemporary Problems* 28:185-212.
- BEVERIDGE, ALBERT J. 1916-1919 *The Life of John Marshall*. 4 vols. Boston: Houghton Mifflin. → Volume 1: *Frontiersman, Soldier, Lawmaker, 1755-1788*. Volume 2: *Politician, Diplomatist, Statesman, 1789-1801*. Volume 3: *Conflict and Construction, 1800-1815*. Volume 4: *The Building of the Nation, 1815-1835*.
- BEYLE, HERMAN C. 1931 *Identification and Analysis of Attribute-cluster-blocs: A Technique for Use in the Investigation of Behavior in Governance*. Univ. of Chicago Press.
- BRECHT, ARNOLD 1959 *Political Theory: The Foundations of Twentieth-century Political Thought*. Princeton Univ. Press.
- CAIRNS, HUNTINGTON 1935 *Law and the Social Sciences*. New York: Harcourt.
- CARDOZO, BENJAMIN N. (1921) 1960 *The Nature of the Judicial Process*. New Haven: Yale Univ. Press.
- CARDOZO, BENJAMIN N. 1924 *The Growth of the Law*. New Haven: Yale Univ. Press.
- CARDOZO, BENJAMIN N. 1928 *The Paradoxes of Legal Science*. New York: Columbia Univ. Press.
- COLE, KENNETH C. 1948 *The Theory of the State as a Sovereign Juristic Person*. *American Political Science Review* 42:16-31.
- CORWIN, EDWARD S. (1919) 1921 *John Marshall and the Constitution: A Chronicle of the Supreme Court*. New Haven: Yale Univ. Press.
- CORWIN, EDWARD S. (1928-1929) 1959 *The "Higher Law" Background of American Constitutional Law*. Ithaca, N.Y.: Cornell Univ. Press.
- DANELSKI, DAVID J. 1961 *The Influence of the Chief Justice in the Decisional Process*. Pages 497-508 in Walter F. Murphy and C. Herman Pritchett (editors), *Courts, Judges, and Politics*. New York: Random House.
- DANELSKI, DAVID J. 1964 *A Supreme Court Justice Is Appointed*. New York: Random House.
- Essays on the Law and Practice of Governmental Administration: A Volume in Honor of Frank J. Goodnow*.



- Edited by Charles Grove Haines and Marshall E. Dimock. 1935. Baltimore: Johns Hopkins Press.
- FAIRMAN, CHARLES 1939 *Mr. Justice Miller and the Supreme Court: 1862-1890*. Cambridge, Mass.: Harvard Univ. Press.
- FRANK, JEROME (1930) 1949 *Law and the Modern Mind*. New York: Coward.
- GOODNOW, FRANK J. (1893) 1903 *Comparative Administrative Law: An Analysis of the Administrative Systems, National and Local, of the United States, England, France and Germany*. 2 vols. New York: Putnam.
- GOODNOW, FRANK J. 1905 *The Principles of the Administrative Law of the United States*. New York: Putnam.
- GOODNOW, FRANK J. 1911 *Social Reform and the Constitution*. New York: Macmillan.
- GRAHAM, HOWARD J. 1938 "The 'Conspiracy Theory' of the Fourteenth Amendment." *Yale Law Journal* 47: 371-403.
- GRANT, J. A. C. 1958 "The Viet Nam Constitution of 1956." *American Political Science Review* 52:437-462.
- GROSSMAN, JOEL B. 1965 *Lawyers and Judges: The ABA and the Politics of Judicial Selection*. New York: Wiley.
- HAINES, CHARLES GROVE 1922 "General Observations on the Effects of Personal, Political, and Economic Influences in the Decisions of Judges." *Illinois Law Review* 17:98-116.
- HAINES, CHARLES GROVE 1930 *The Revival of Natural Law Concepts: A Study of the Establishment and of the Interpretation of Limits on Legislatures With Special Reference to the Development of Certain Phases of American Constitutional Law*. Cambridge, Mass.: Harvard Univ. Press.
- HART, H. L. A. 1961 *The Concept of Law*. Oxford: Clarendon.
- HAYAKAWA, TAKEO (1962) 1964 "Civil Liberties in the Japanese Supreme Court." Pages 325-334 in Glendon A. Schubert (editor), *Judicial Behavior: A Reader in Theory and Research*. Chicago: Rand McNally. → A partial reprinting of "Legal Science and Judicial Behavior, With Particular Reference to Civil Liberties in the Japanese Supreme Court," Number 2, pages 1-27 in the *Kobe University Law Review*.
- HYNEMAN, CHARLES S. 1959 *The Study of Politics: The Present State of American Political Science*. Urbana: Univ. of Illinois Press.
- JACOB, HERBERT 1965 *Justice in America: Courts, Lawyers, and the Judicial Process*. Boston: Little.
- JOHNSON, RICHARD M. 1965 "Separation of Church and State: The Dynamics of Supreme Court Decision-making." Ph.D. dissertation, Univ. of Illinois.
- KELSEN, HANS (1945) 1961 *General Theory of Law and State*. New York: Russell.
- KORT, FRED 1957 "Predicting Supreme Court Decisions Mathematically: A Quantitative Analysis of the 'Right to Counsel' Cases." *American Political Science Review* 51:1-12.
- LESSWELL, HAROLD D. 1948 *Power and Personality*. New York: Norton.
- LESSWELL, HAROLD D.; and McDUGAL, MYRES S. 1943 "Legal Education and Public Policy: Professional Training in the Public Interest." *Yale Law Journal* 52:203-295.
- LATHAM, EARL (1952) 1965 *The Group Basis of Politics: A Study in Basing-point Legislation*. New York: Octagon Books.
- LLEWELLYN, KARL N. (1934) 1935 "The Constitution as an Institution." Pages 277-322 in *Legal Essays in Tribute to Orrin Kip McMurray*. Edited by Max Radin and A. M. Kidd. Berkeley: Univ. of California Press.
- MCCLOSKEY, ROBERT G. 1960 *The American Supreme Court*. Univ. of Chicago Press.
- MAGRATH, C. PETER 1963 *Morrison R. Waite: The Triumph of Character*. New York: Macmillan.
- MASON, ALPHEUS T. 1956 *Harlan Fiske Stone: Pillar of the Law*. New York: Viking.
- MENDELSON, WALLACE 1963 "The Neo-behavioral Approach to the Judicial Process: A Critique." *American Political Science Review* 57:593-603.
- MERRIAM, CHARLES E. 1921 "The Present State of the Study of Politics." *American Political Science Review* 15:173-185.
- MERRIAM, CHARLES E. 1926 "Progress in Political Research." *American Political Science Review* 20:1-13.
- MITAU, THEODORE G. 1965 "What Has Happened to the Study of State Public Law by Political Scientists? A Note on Achievements and Lacunae in the Study of State Constitutional Law Since 1950." *Journal of Public Law* 14:90-104.
- MOTT, RODNEY L.; ALBRIGHT, S. D.; and SEMMERLING, H. R. 1933 "Judicial Personnel." *American Academy of Political and Social Science, Annals* 167:143-155.
- MURPHY, WALTER F. 1964 *Elements of Judicial Strategy*. Univ. of Chicago Press.
- PATRICK, GORDON 1957 "The Impact of a Court Decision: Aftermath of the McCollum Case." *Journal of Public Law* 6:455-464.
- PELTASON, JACK W. 1953 "A Political Science of Public Law." *Southwestern Social Science Quarterly* 34, no. 2:51-56.
- PELTASON, JACK W. 1955 *Federal Courts in the Political Process*. Garden City, N.Y.: Doubleday.
- PELTASON, JACK W. 1961 *Fifty-eight Lonely Men: Southern Federal Judges and School Desegregation*. Introduction by Paul Douglas. New York: Harcourt.
- POWELL, THOMAS REED 1918 "The Logic and Rhetoric of Constitutional Law." *Journal of Philosophy, Psychology and Scientific Methods* 15:645-658.
- PRITCHETT, C. HERMAN 1941 "Divisions of Opinion Among Justices of the U.S. Supreme Court: 1939-1941." *American Political Science Review* 35:890-898.
- PRITCHETT, C. HERMAN (1948) 1963 *The Roosevelt Court: A Study in Judicial Politics and Values, 1937-1947*. New York: Octagon Books.
- PRITCHETT, C. HERMAN 1954 *Civil Liberties and the Vinson Court*. Univ. of Chicago Press.
- SCHUBERT, GLENDON 1958 "The Study of Judicial Decision-making as an Aspect of Political Behavior." *American Political Science Review* 52:1007-1025.
- SCHUBERT, GLENDON 1963 "Behavioral Research in Public Law: Bibliographical Essay." *American Political Science Review* 57:433-445.
- SCHUBERT, GLENDON 1965 *The Judicial Mind: The Attitudes and Ideologies of Supreme Court Justices, 1946-1963*. Evanston, Ill.: Northwestern Univ. Press.
- SHEPARD, MAX 1939 "Law and Obedience." *American Political Science Review* 33:783-810.
- SHERWOOD, FOSTER H. 1958 "The Role of Public Law in Political Science." Pages 86-96 in Roland Young (editor), *Approaches to the Study of Politics*. Evanston, Ill.: Northwestern Univ. Press.
- SOMIT, ALBERT; and TANENHAUS, JOSEPH 1964 *American Political Science: A Profile of a Discipline*. New York: Atherton.

- SWISHER, CARL B. (1935) 1961 *Roger B. Taney*. Hamden, Conn.: Shoe String Press.
- SWISHER, CARL B. (1943) 1954 *American Constitutional Development*. 2d ed. Boston: Houghton Mifflin.
- SWISHER, CARL B. 1946 Research in Public Law: Report of the Panel on Public Law. *American Political Science Review* 40:552-562.
- TRUMAN, DAVID B. (1951) 1962 *The Governmental Process: Political Interests and Public Opinion*. New York: Knopf.
- VOSE, CLEMENT E. 1958 Litigation as a Form of Pressure Group Activity. *American Academy of Political and Social Science, Annals* 319:20-31.
- VOSE, CLEMENT E. 1959 *Caucasians Only: The Supreme Court, the NAACP, and the Restrictive Covenant Cases*. Berkeley: Univ. of California Press.
- WALZ, GUSTAV ADOLPH 1934 Public Law. Volume 12, pages 657-659 in *Encyclopaedia of the Social Sciences*. New York: Macmillan.
- WESTIN, ALAN F. 1958 *The Anatomy of a Constitutional Law Case: Youngstown Sheet and Tube Co. v. Sawyer: The Steel Seizure Decision*. New York: Macmillan.
- WILLOUGHBY, W. W. 1896 *An Examination of the Nature of the State: A Study in Political Philosophy*. New York: Macmillan.
- WILLOUGHBY, W. W. 1924 *The Fundamental Concepts of Public Law*. New York: Macmillan.
- WRIGHT, BENJAMIN F. 1938 *The Contract Clause of the Constitution*. Cambridge, Mass.: Harvard Univ. Press.

## II

## COMPARATIVE PUBLIC LAW

Comparative public law is that academic discipline which investigates problems of public law by use of the comparative method.

## Scope, aims, and potentialities

Constitutional law, administrative law, criminal law, tax law, labor law, welfare law, and antitrust law, as well as the law relating to civil and criminal procedure, are now considered to be "public law" even in countries which traditionally did not differentiate between public and private law along the lines followed by legal theory in continental Europe. It can hardly be denied that these fields deal with quite disparate matters involving great varieties of social, political, and organizational problems, and have little in common. Any attempt to define public law in terms of a self-contained category suffers from the ambiguities inherent in the dichotomy of "public" and "private" or in such distinction as that between the interests of the individual and the commonweal. Historically the bipartition found its classical formulation in Ulpian's statement: *publicum ius est quod ad statum rei Romanae spectat, privatum quod ad singulorum utilitatem* ("Public law is that which focuses upon the status of the Roman commonwealth, while private law pertains to the interests of individuals": *Digest* 1, 1, 1, 2). It played a

shadowy role during the Middle Ages and gained new currency during the seventeenth century with the emergence of the modern state. During the nineteenth century, when the distinction between private and public law had its heyday, it paralleled the supposed dualism between state and society. Currently, however, this antinomy has been abandoned because it rests on a contraposition of the state, conceived as an independent entity, to society, likewise conceived as an independent entity, a theory closely linked to nineteenth-century European liberal doctrine. Yet the views regarding the antithesis between public and private law are significant from a comparative point of view, since they suggest varying notions about the relationships between the individual and collective spheres in time and space.

No less perplexing than the borderline between comparative public law and comparative private law is the line of demarcation which separates comparative public law from comparative studies in the fields of sociology of law and political science. Although at the turn of the century links between law and the newly emerging disciplines of political science and sociology were still rather close, the subsequent emphasis of political science and sociology on "factual" inquiries rather than on "normative" data led to an increasing divergence. The legal perspectives in early political science are exemplified by the impact of the German *allgemeine Staatslehre* on Burgess (1890-1891) and of European administrative law on Goodnow (1893); Max Weber's work in sociology also reflects his legal background.

Conversely, modern legal scholarship, stimulated by contemporary schools of jurisprudence, such as Roscoe Pound's sociological school, the legal realists, the policy-oriented approach of Lasswell and McDougal, and the psychoanalytical interpretation, has embarked on a methodological reorientation and endeavors to utilize the full scope of historical, economic, sociological, and behaviorist techniques. As a result, disciplinary boundaries have become of secondary importance. Perhaps it can be said that the main distinction between comparative public law and other comparative social sciences is the vantage point from which phenomena are selected and viewed.

Another set of boundaries of the scope of comparative public law results from restrictions inherent in the comparative method itself. Except for cases of mere cataloging or sheer contrast, fruitful comparison is predicated on some similarity of political or economic organization, technological advance, or value scales.



The difficulties attending upon the requisite comparability are least pronounced with respect to institutional techniques devised to respond to common problems and needs, e.g., social security arrangements established to alleviate the effects of economic dependency flowing from an industrialized society. In such areas, public law comparison is similar to comparative private law studies and can benefit from their methodology: functional analysis applied to the solution of specific policy problems. In fields where political habits, institutions, and value judgments play a dominant part, comparative studies must properly assess the role of these factors. Needless to say, the elements of the requisite comparability vary greatly with the nature, range, and purposes of the comparative inquiry.

The aims of comparative public law, like the purposes of comparative law in general, are both cognitive and normative. In this respect it has the protean aspects of all legal studies. In the cognitive perspective comparative public law parallels comparative empirical research in the social sciences. In the normative perspective it manifests the specific attributes of all legal research, i.e., it is directed toward law reform. In designating comparative legal studies as primarily cognitive, one does not, of course, exclude normative elements altogether, since a better understanding of the way in which legal institutions operate may in turn influence their acceptance or rejection. From this vantage point comparative public law possesses major educational significance.

The practical value of comparative public law has traditionally been emphasized in two areas: negotiation and application of multinational treaties, and law reform. A modern example of the importance of comparative public law in the interpretation of multinational agreements is furnished by the construction of the term "misuse of power" as a ground for the annulment of administrative acts (see Coal and Steel Community Treaty, art. 33; European Economic Community Treaty, art. 173; Euratom Treaty, art. 146). In arriving at the proper meaning, the Court of Justice for the European Economic Community, basing its holding on the comparative discussions of the advocate general, searched for a common core in the administrative laws of the community countries.

In the field of law reform, the examples of the use of comparative public law are countless. It can be said without exaggeration that since the beginning of the twentieth century no major constitutional revision or introduction of important new legal institutions has taken place without extensive

comparative studies. The establishment of separate courts for deciding constitutional questions in several European countries, the creation of the social security system in the United States, and the development of the Mexican writ of *amparo* in analogy to habeas corpus provide good illustrations.

Although it has been suggested by pioneers in legal comparison, such as H. C. Gutteridge (1946), that the extent to which the comparative process may yield valuable results is open to doubt when problems of political significance are under consideration, experience seems to have shown that given proper awareness of the ideological ramifications, even the transplantation of major political institutions is possible. Gneist's studies of British local self-government (1857-1860) furnished a powerful impetus toward a restructuring of Prussian municipal organization. Moreover, in some Western countries, recent interludes of departure from established principles, especially in the area of civil liberties, have led to re-examinations of the legal systems of these countries intended to test their consonance with accepted values and goals of the Western tradition. Certainly much of modern German and Italian legal literature is the outgrowth of such efforts in comparative law.

### History

The growth of comparative public law came in the wake of the emergence of the modern nation-state, as an aftermath of the rise of different social and political patterns within the orbit of the Western world, and as a result of the acceptance of comparison as a tool of scholarly inquiry. The work of Jean Bodin at the end of the sixteenth century may be considered as the starting point and an early masterpiece of the modern discipline [see *BODIN*]. His analysis of the state and the role of government was buttressed by an impressive and far-flung comparison of the existing realities of his time seen in the light of legal principles derived from Roman law. The works of Leibniz and Vico in the next century and a half reflect the effort to arrive at the essence of legal authority and government on the basis of historical and comparative insights. [See the *biography* of Vico.]

In contrast, the writers of the subsequent period emphasized the effects of environmental factors as causes of diversity and attempted comparisons on that basis. Undoubtedly Montesquieu's *L'esprit des lois* (1748) heralded a new era in comparative constitutional law and made an indelible impression on his time [see *MONTESQUIEU*]. Although Montesquieu's comparison of the French and English constitutions was tendentious and lop-

sided, it proved the value and inherent political potentialities of such studies. Accordingly, comparative surveys and critiques in Montesquieu's fashion were widely used by authors following him. The most notable studies of that type are Robertson's "View of the Progress of Society in Europe, With Respect to Interior Government, Laws and Manners," which was an essay preliminary to his comprehensive *History of the Reign of the Emperor Charles the Fifth* (1769), and Filangieri's *Science of Legislation* (1780-1785). Robertson's essay was familiar to and quoted by members of the Constitutional Convention of 1787, especially in conjunction with the discussion of judicial review, and Filangieri's treatise is known to have elicited Franklin's admiration. Moreover, Filangieri, whose chapter on the divisions of criminal law demonstrably relied on Blackstone's genius for systematization, exerted a far-reaching impact on the early continental European codifications made not long after his death.

During the nineteenth century, comparative public law in Europe, and particularly in Germany, shifted its principal focuses to criminal law and administrative law. In criminal law it was particularly Anton Mittermaier who, following the precepts of his teacher Ludwig Feuerbach, most fruitfully employed the comparative inquiry which gained him international repute. His penchant for comparison was manifested in all of his literary efforts but was most fruitfully employed in his endeavors toward a far-reaching reform of criminal procedure based on insights gained from French, English, and American practice. His predilection for comparison caused his rejection of the doctrinal-conceptualistic approach then in vogue with many of his contemporaries. In 1829, Mittermaier and Heinrich Albert Zachariae founded the *Kritische Zeitschrift für Rechtswissenschaft und Gesetzgebung des Auslandes*, which was the first periodical in the world devoted to the study of comparative law. The works of Gneist in administrative law and their importance have already been mentioned.

During the middle of the nineteenth century, scope and methodology of comparative law, including comparative public law, became a matter of academic discussion and controversy. Emerico Amari set off that debate with his *Critica di una scienza delle legislazioni comparate* (1857). Although the work itself was soon forgotten and properly appreciated only much later, its initial impact persisted. The questions whether comparative law was a subject of its own or merely a method and whether it should be treated from a philosophical or historical-sociological point of

view especially became and remained matters of much discussion. In Germany, the work of Joseph Kohler and his school concentrated on the ethnological-cultural aspects of the field. But the theoretical disputes were soon pushed into the background by the increasing institutionalization of the field: the establishment of societies for the study of comparative legislation, the creation of institutes of comparative law, and the publication of periodicals devoted exclusively to comparative law and comparative legislation.

This development started in France in 1869, with the formation of the Société de Législation Comparée. In the year of its establishment the society commenced publication of its celebrated *Bulletin*. Even prior to that date the French jurist Foelix published the *Revue étrangère de législation et d'économie politique*, which, under various titles, appeared between 1833 and 1849. It counted on its roster of distinguished foreign collaborators such eminent American jurists as Joseph Story and Francis Lieber. Likewise, in 1869 the *Revue de droit international et de législation comparée*, printed in Belgium, was established by Asser, Rolin-Jacquemyns, and Westlake. The German *Zeitschrift für vergleichende Rechtswissenschaft*, edited by Bernhöft, Cohn, and, after the first two volumes, Kohler, followed in 1878. In England the Society of Comparative Legislation was formed in 1895 and commenced to publish the *Journal of Comparative Legislation and International Law* in 1899.

During the twentieth century, especially in the wake of World War I, the study of comparative law has become much expanded and intensified. Institutes of comparative law have been established in many universities of the world. During this period comparative public law has emerged as a separate branch of comparative legal studies. In Germany this development found an outward manifestation in the establishment, by the Kaiser Wilhelm Gesellschaft zur Förderung der Wissenschaften, of a separate Institut für Ausländisches Öffentliches Recht und Völkerrecht (now the Max-Planck-Institut in Heidelberg) alongside the Institut für Ausländisches und Internationales Privatrecht in 1926. In France the *Revue du droit public et de la science politique en France et à l'étranger*, established in 1894, devoted special attention to comparative constitutional and administrative law. In the United States, the University of California instituted a course of lectures in comparative public law, given in 1917 by Ludwik Ehrlich; and the Harvard Law School appointed Joseph Redlich, a specialist in comparative public law, as its first



professor of comparative law in 1926. The increased importance of comparative studies in the various branches of public law led to the gradual emergence of a number of relatively independent subfields.

### Major areas of comparative public law

**Comparative constitutional law.** Comparative constitutional law is probably the oldest branch of comparative public law. The notions of popular sovereignty and constitutionalism, which demanded that all powers of government be fixed in a fundamental charter, determining the organization and powers of the authorities, gave a powerful impetus to studies of that type [see SOVEREIGNTY]. Among English-speaking people they were manifested in the utopian literature, particularly in Harrington's *Commonwealth of Oceana*. During Cromwell's regime its author actually undertook a comparative study of constitutions, and his ideas exerted a demonstrable impact both on the constitution makers in the United States and on the political conceptions of Montesquieu. The American experiment, in turn, had its repercussions on the constitutional ideas of the French Revolution. During the middle of the nineteenth century the quest for written constitutions found a fertile ground in Germany. The acquisition of independence by the Spanish colonies in the New World was followed by the adoption of written constitutions by the newly formed states. The constitution of Argentina of 1853 was directly modeled after the constitution of the United States, although other constitutional arrangements were not neglected, and the constitutional decisions of the Supreme Court of the United States were frequently relied upon and quoted in the adjudications of the Argentine Supreme Court. The British North America Act of 1867 and the Australia Constitution Act of 1900, which, on a federal basis, form the governmental framework of Canada and Australia, respectively, likewise were the product of comparative studies and in turn provided a strong motivation for further comparative researches. In modern days many new nations formed in other parts of the globe have been confronted with the problem of the extent to which they should borrow from existing constitutional models [see CONSTITUTIONS AND CONSTITUTIONALISM]. In addition, the creation of supranational communities in Europe and elsewhere has given new significance to comparative studies of federal structures [see FEDERALISM]. Other matters which have attracted particular attention from students of comparative constitu-

tional law are the constitutional recognition and judicial protection of human rights and the experiences with the judicial control of the constitutionality of legislation. [See CONSTITUTIONAL LAW.]

**Comparative administrative law.** Comparative administrative law is a more recent branch of comparative law. It has taken its subject matter from an assessment of the Council of State of France, which during the second half of the nineteenth century developed into the supreme administrative tribunal of that nation and succeeded in building up administrative law as a separate body of legal principles governing administrative action. The success of this court inspired Otto Mayer in the late nineteenth century to develop comparative administrative law as a special field of comparative inquiry. Scholars in other countries, especially Frank Goodnow in the United States, followed in his footsteps [see GOODNOW]. Comparison in administrative law has focused on three major issues concerning the relationship between the citizen and public authorities that were prompted by the ever-increasing scope of the activities of the modern state: (1) the judicial control of the legality of administrative action, (2) the attribution of such control to separate administrative tribunals instead of to the ordinary courts, and (3) the organization of such tribunals. Contemporary critics maintain that the ordinary courts in Great Britain and the United States have failed to exercise adequate supervision over legality and abuse of discretion in administrative action and to advocate the use of Continental models. Against them, however, it has been argued that bifurcation of the administration of justice results in unnecessary conflicts between jurisdictions and in a break in the uniformity of the legal system. Moreover, this different attitude regarding the scope of judicial control stems from different ideas about the range of policy judgments properly left to the bureaucracy and from different emphases placed on policy-making stage and policy execution. [See ADMINISTRATIVE LAW.]

**Comparative criminal law.** Research in comparative criminal law received its original impetus from the publication of Beccaria's *An Essay on Crimes and Punishments* (1764) and the resulting reform movement [see BECCARIA]. It prompted the announcement of a prize for the best work proposing a plan for complete and detailed legislation covering substantive criminal law and criminal procedure, including the process of proof. Voltaire took a leading part in that venture. Since those days the problem of the proper treatment of crimi-

nals has been much discussed both internationally and comparatively in Europe and elsewhere. The University of Paris established a special chair for comparative criminal law as early as 1846, Joseph Louis Ortolan being its first incumbent. In England the "Fourth Report" of the Royal Commissioners of Criminal Law (Great Britain . . . 1834-1845) noted expressly that their reform proposals were based on a study of the contemporary criminal codes in Europe and the United States and the writings of foreign jurists, and in Germany the preparation of a new penal code induced the German Ministry of Justice to sponsor the systematic comparison of foreign criminal legislation. A 16-volume treatise entitled *Vergleichende Darstellung des deutschen und ausländischen Strafrechts* (1905-1909) was produced, a work which has been hailed as a landmark in the history of comparative studies. Today such institutions as the indeterminate sentence, preventive measures, and adult and juvenile authorities have been subjects of elaborate comparative study. The whole range of the causes and control of delinquency is particularly appropriate for such types of inquiry, and criminological research has become truly international in scope. [See CRIMINAL LAW.]

**Comparative civil procedure.** The range of comparative legal studies has also been extended to the field of civil procedure and related subjects, such as bankruptcy. The roles of the judge and of the parties in a civil trial, the rules of evidence and testimonial privileges, and the effect of adjudications are the subject of numerous modern comparative studies in various countries.

**Comparative labor and social security law.** The process of industrialization has given rise to many institutions which have developed in a parallel fashion in many countries, for example, collective bargaining and social insurance systems. In many instances the legislation of certain pioneer nations in the field has been transplanted and adapted by other nations. Much of the comparative work is done through international organizations such as the International Labor Office or international associations such as the International Association for Social Security.

**Comparative tax and antitrust law.** The tremendous growth of the functions of the modern state has been accompanied by a parallel increase in its need for revenues. As a result, the law of taxation has become one of the most important branches of public law, and thus has become the subject of a vast number of comparative studies. Moreover, the harmonization of tax laws is one of

the difficult tasks to be faced by the new supranational organizations. Finally, it should be noted that the current efforts to create economies of scale and the policy of relying on untrammelled market forces as a means of economic progress have led to the growth of comparative antitrust law as a new and extended discipline [see ANTITRUST LEGISLATION; TAXATION].

STEFAN A. RIESENFELD  
AND GERHARD CASPER

[See also CONFLICT OF LAWS; LEGAL SYSTEMS, article on COMPARATIVE LAW and LEGAL SYSTEMS. Other relevant material may be found in INTERNATIONAL LAW; POLITICS, COMPARATIVE; and in the articles listed in the guide to the reader under LAW.]

#### BIBLIOGRAPHY

##### BOOKS AND ARTICLES

- AMARI, EMERICO 1857 *Critica di una scienza delle legislazioni comparate*. Genoa (Italy): Sordomuti.
- BECCARIA, CESARE BONESANA (1764) 1953 *An Essay on Crimes and Punishments*. Stanford, Calif.: Academic Reprints. → First published in Italian under the title *Dei delitti e delle pene*. A paperback edition was published in 1963 by Bobbs-Merrill.
- BERLIN, MAX-PLANCK-INSTITUT FÜR AUSLÄNDISCHES ÖFFENTLICHES RECHT UND VÖLKERRECHT 1962 *Verfassungsgerichtsbarkeit in der Gegenwart: Länderberichte und Rechtsvergleichung; Internationales Kolloquium. Beiträge zum ausländischen öffentlichen Recht und Völkerrecht*, No. 36. Cologne (Germany): Heymann.
- BOWIE, ROBERT R.; and FRIEDRICH, CARL J. (editors) 1954 *Studies in Federalism*. Boston: Little.
- BURGESS, JOHN W. (1890-1891) 1902 *Political Science and Comparative Constitutional Law*. 2 vols. Boston: Ginn. → Volume 1: *Sovereignty and Liberty*. Volume 2: *Government*.
- EHRlich, LUDWIK 1921 *Comparative Public Law and the Fundamentals of Its Study*. *Columbia Law Review* 21:623-646.
- FILANGIERI, GAETANO (1780-1785) 1806 *The Science of Legislation*. 2 vols. London: Ostell. → First published as *La scienza della legislazione*.
- FONTANA, GIOVANNI (1938) 1954 *Introduzione al diritto pubblico comparato*. Florence (Italy): CYA.
- FRIEDRICH, CARL J. (1937) 1950 *Constitutional Government and Democracy: Theory and Practice in Europe and America*. Rev. ed. Boston: Ginn. → First published as *Constitutional Government and Politics: Nature and Development*.
- GNEIST, RUDOLF (1857-1860) 1871-1884 *Das heutige englische Verfassungs- und Verwaltungsrecht*. 3 parts, 3d ed. Berlin: Springer. → Volume 1: *Das englische Verwaltungsrecht der Gegenwart in Vergleichung mit den deutschen Verwaltungssystemen*, 1883-1884. 2 parts. Volume 2: *Selfgovernment: Communalverwaltung und Verwaltungsgerichte in England*, 1871.
- GOODNOW, FRANK J. (1893) 1903 *Comparative Administrative Law: An Analysis of the Administrative Systems, National and Local, of the United States*,



## PUBLIC OPINION

England, France and Germany. 2 vols. New York: Putnam.

GREAT BRITAIN, COMMISSIONERS ON CRIMINAL LAW 1834-1845 First— [Eighth] Report From His Majesty's Commissioners on Criminal Law. 2 vols. London: Printed by W. Clowes and Sons for His Majesty. . . .  
→ See especially the "Fourth Report."

GUTTERIDGE, HAROLD C. (1946) 1949 *Comparative Law: An Introduction to the Comparative Method of Legal Study and Research*. 2d ed. Cambridge Univ. Press.

HALL, JEROME 1963 *Comparative Law and Social Theory*. Baton Rouge: Louisiana State Univ. Press.

KAISER, J. H. et al. 1964 *Vergleichung im öffentlichen Recht. Zeitschrift für ausländisches öffentliches Recht und Völkerrecht* 24:391-471.

KÜLZ, HELMUT R.; and NAUMANN, RICHARD (editors) 1963 *Staatsbürger und Staatsgewalt: Verwaltungsrecht und Verwaltungsgeschichte in Geschichte und Gegenwart*. Vol. 1. Karlsruhe (Germany): Müller.

LEIBNIZ, GOTTFRIED W. (1667) 1768 *Nova methodus discendae docendaeque jurisprudentiae*. Volume 4, part 3, pages 169-624 in Gottfried W. Leibniz, *Opera omnia*. Geneva: Fratres de Tournes.

LINARES QUINTANA, SEGUNDO V. 1953— *Tratado de la ciencia del derecho constitucional argentino y comparado*. Vol. 1—. Buenos Aires: Alfa. → Nine volumes were published as of 1966.

MONTESQUIEU, CHARLES (1748) 1962 *The Spirit of the Laws*. 2 vols. New York: Hafner. → First published in French.

RIESENFELD, STEFAN 1938 *The French System of Administrative Justice: A Model for America?* *Boston University Law Review* 18:48-82, 400-432, 715-748.

ROBERTSON, WILLIAM (1769) 1916 *The History of the Reign of Emperor Charles the Fifth*. 3 vols. Philadelphia: Lippincott.

*Vergleichende Darstellung des deutschen und ausländischen Strafrechts*. 16 vols. 1905-1909 Berlin: Liebmann.

VICO, GIOVANNI BATTISTA (1720) 1858 *Dell'unico principio e dell'unico fine del diritto universale*. Milan (Italy): Agnelli. → First published in Latin.

## SERIES AND JOURNALS

HARVARD UNIVERSITY, INTERNATIONAL PROGRAM IN TAXATION *World Tax Series*. → Published since 1957.

*International and Comparative Law Quarterly*. → Published since 1952. Supersedes the *Journal of Comparative Legislation* and the *International Law Quarterly*.

INTERNATIONAL BUREAU OF FISCAL DOCUMENTATION *Publications*. → Published since 1951. A series of short studies on international and comparative fiscal law.

INTERNATIONAL SOCIAL SECURITY ASSOCIATION *Bulletin*. → Published since 1948.

*Jahrbuch des öffentlichen Rechts der Gegenwart* New Series. → Published since 1951.

*Kritische Zeitschrift für Rechtswissenschaft und Gesetzgebung des Auslandes*. → Published between 1829 and 1856.

*Revue de droit international et de législation comparée*. → Published from 1869 to 1940.

*Revue du droit public et de la science politique en France et à l'étranger*. → Published since 1894.

*Zeitschrift für ausländisches öffentliches Recht und Völkerrecht*. → Published since 1929.

*Zeitschrift für vergleichende Rechtswissenschaft*. → Published since 1878.

## I. INTRODUCTION

W. Phillips Davison

## II. POLITICAL OPINION

Avery Leiserson

## INTRODUCTION

There is no generally accepted definition of "public opinion." Nevertheless, the term has been employed with increasing frequency since it came into popular usage at the time of the French Revolution, when Louis XVI's finance minister, Jacques Necker, referred to public opinion as governing the behavior of investors in the Paris money market. Later efforts to define the term precisely have led to such expressions of frustration as: "Public opinion is not the name of a something, but a classification of a number of somethings." (See Childs 1965, pp. 12-28, for some fifty different definitions.)

In spite of differences in definition, students of public opinion generally agree at least that it is a collection of individual opinions on an issue of public interest, and they usually note that these opinions can exercise influence over individual behavior, group behavior, and government policy. Because public opinion is acknowledged to play a role in several diverse areas, leading writers on the subject have included sociologists (Tönnies 1887; Lazarsfeld et al. 1944; Albig 1956), political theorists (Bryce 1888; Lasswell 1927; Lippmann 1922), social psychologists (Allport 1937; Cantril 1966), and historians (Bauer 1929). Those who are engaged in manipulating public opinion have also made important contributions: for example, politicians (Lenin 1929) and public-relations specialists (Bernays 1923). Differences in definition and approach can be accounted for largely by the differing interests of various categories of students and practitioners.

The principal approaches to the study of public opinion may be divided into four partially overlapping categories: quantitative measurement of opinion distributions; investigation of the internal relationships among the individual opinions that make up public opinion on an issue; description or analysis of the political role of public opinion; and study both of the communication media that disseminate the ideas on which opinions are based and of the uses that propagandists and other manipulators make of these media.

Some researchers have simultaneously contributed to knowledge in several of these categories. An investigation in Erie County, Ohio, of the 1940

U.S. presidential election campaign not only provided statistical measurements of voting intentions over time but also explored the influence of group membership on individual opinions and evaluated the impact of mass communications on the outcome of the election (Lazarsfeld et al. 1944). Another example may be taken from a study of Norwegian opinions on a number of foreign policy issues between 1959 and 1964 (Galtung 1964). In this case the researcher was able to contribute to the body of theory regarding political behavior in a democracy by analyzing the relationship between foreign policy attitudes and social position while also taking into account the role of group affiliation and the communication structure of the society.

### Quantitative measurement of opinions

A definition offered by a specialist in polling is suggestive of the approach to public-opinion research of those who are primarily interested in measurement: "Public opinion consists of people's reactions to definitely worded statements and questions under interview conditions" (Warner 1939, p. 377). Those focusing on measurement usually investigate such questions as the following: How widely (and, sometimes, how intensely) is a given opinion held? In which geographic, religious, ethnic, or socioeconomic sectors is the opinion encountered most frequently? With what other opinions is it most closely associated?

Measurement of opinions based on polling representative samples of larger populations came into prominence in the United States following the presidential election of 1936. At that time George Gallup, Elmo Roper, and Archibald Crossley, then the most prominent exponents of the sample-survey method, correctly predicted the outcome, while the *Literary Digest*, relying on nearly 2.5 million unrepresentative straw ballots, was off by nearly twenty percentage points. Use of the sample-survey method spread rapidly thereafter and was scarcely affected by the failure of the principal polling organizations to pick the winner in the presidential election of 1948. This failure did, however, lead to important improvements in polling methods. (For a succinct description of polling procedures in recent election surveys, see Perry 1960.)

By 1965 public-opinion polling had spread throughout the world. The World Association for Public Opinion Research had members from more than forty countries, and numerous polling organizations were reported to be working in communist and other countries that were not represented in

the association's membership. The New York-based International Research Associates, headed by Elmo C. Wilson, had branch offices or affiliates in 34 countries, and the Gallup poll counted 32 affiliates and conducted frequent polls in nearly fifty countries. In the United States several hundred survey organizations existed on a national, state, or local level, with university research bureaus accounting for a substantial number of these.

Data gathered in the course of numerous surveys throughout the world are centralized in a number of "data banks," the oldest of which is the Roper Public Opinion Research Center at Williams College. The Council of Social Science Data Archives, a cooperative organization of American university and non-profit research groups, helps to promote the exchange of these basic data for purposes of secondary analysis. Columbia University's Bureau of Applied Social Research serves as a secretariat for the council. A selection of poll results from the United States appears in each issue of the *Public Opinion Quarterly*, and selected results from throughout the world are carried in *Polls*, a magazine published by Systemen Keesing in Amsterdam.

Quantitative studies have led to numerous generalizations about public opinion, most of which, however, do not hold for every time and place. One is that large numbers of people pay surprisingly little attention to political personalities and issues, even when these are featured by the mass media. For instance, in 1964 one out of four adult Americans did not know that there was a communist government on mainland China (Michigan, University of . . . 1964). Polls taken since World War II in the United States have consistently shown that large proportions of the respondents were unaware of such crises as those of 1959 and 1961 in Berlin, or of 1955 in the Formosa Straits. As of 1959, 22 per cent of adult Americans said that they had not heard or read anything about Fidel Castro (Erskine 1963, pp. 661-662). Surveys conducted in western European countries have disclosed similar results: more people are familiar with the names of leading sports and entertainment figures than with all but the most prominent politicians. As of 1961, 95 per cent of rural Brazilians were unable to identify the president of the United States (Institute for International Social Research 1961, p. 3).

Stating the same proposition in another way, one can say that relatively small numbers of people regularly show a serious concern for public affairs. This has led some students to distinguish between



the "general public," the "attentive public" (which is at least aware of important issues), and the "informed public," which participates in discussion of the issues (Almond [1950] 1960, p. 138). It has also led to the polling of "elites," which are variously defined as being composed of those with a high degree of wealth, education, prominence, or influence. For example, numerous polls have been conducted using samples of legislators, businessmen, or those listed in *Who's Who*.

The other side of the coin is that people are likely to be most concerned with matters that they see as affecting them directly. A survey of the principal worries of adult Americans found that 80 per cent of the respondents answered solely in terms of personal and family problems; only 6 to 8 per cent mentioned national or world problems, including war, as being among the things they worried about most, in spite of the fact that the survey was conducted during a period when there was an active atomic-arms race with the Soviet Union (Stouffer 1955). Another study, using cross sections in more than twenty countries throughout the world—including several communist countries and developing countries—found a similar pattern of concerns, although with important national variations (Cantril 1966).

Nor is it a simple matter to raise the level of information about public issues (Hyman & Sheatsley 1947). Repeated experiments and observations in several countries have indicated that people have a remarkable ability to ignore easily available facts when these facts are of little interest to them. Merely increasing the amount of information available will not necessarily increase public knowledge, although this generalization probably will not hold true in developing areas where there is a strong unfulfilled demand for more mass media. What seems to be the case is that a proportion of each population, ranging from a very small group up to about two-thirds of the adults, experiences a need for at least some information on matters of public concern. Once the needs of these people have been satisfied, further information flows over the population like water over a saturated sponge.

Opinion measurement has also disclosed strong correlations between educational, religious, geographic, socioeconomic, and ethnic factors, on the one hand, and the opinions that people hold on political subjects, on the other. Indeed, in several areas of the United States it has been found that a person's voting behavior can be predicted with considerable reliability if information is available about his socioeconomic status, his place of residence (urban or rural), and his religious or ethnic background (Lazarsfeld et al. 1944, p. 26). In

general, political interest and activity are greater among upper socioeconomic and educational levels, men, middle-aged groups, and urban residents than among lower socioeconomic and educational levels, women, older and younger adults, and rural residents (*ibid.*, pp. 42–51). A number of surveys have also found a relationship between political opinions and personality factors—although this relationship has usually been weaker and more difficult to document satisfactorily.

A final example of the kind of generalization that can be developed through quantitative measurement is that people tend to adjust their opinions to conform to the situation in which they find themselves. During racial integration of schools in the United States, for instance, public opinion in the areas affected tended to become more favorable toward integration *following*, rather than before, the action of school authorities to admit Negroes to formerly all-white schools (Hyman & Sheatsley 1964). Similarly, the attitudes that the people of two nations have toward each other are likely to result from the state of relations between their governments, as well as being a cause of good or bad relations (Buchanan & Cantril 1953).

The use of public-opinion studies. The rapid spread of public-opinion measurement around the world is a reflection of the number of uses to which it can be put. Governments have increasingly found surveys to be useful tools for guiding their public-information and propaganda programs and occasionally for helping in the formulation of other kinds of policies. The U.S. Department of Agriculture was one of the first government agencies to sponsor systematic and large-scale surveys. It was followed by many other federal bodies, including the U.S. Information Agency, which has conducted opinion research in all parts of the world in which such activities are permitted. Public-opinion studies made abroad following the orbiting of the first Soviet satellite contributed to the decision of the United States to speed up its own space program. Similarly, governments of many other countries now either commission polls on questions of domestic and foreign policy or pay close attention to surveys made under private auspices. France, West Germany, and Japan are among the most frequent users of this kind of research, and there are few major countries that have not commissioned one or more surveys for policy purposes. India and Indonesia are among the newer nations that have occasionally turned to opinion measurement (Free 1967).

Individual politicians, as well as governments, have found polls useful. Opinion surveys can give

them a rough estimate of their chances of election and can also help them gauge the salience of various issues with the voters and evaluate the effectiveness of their campaign propaganda. Once in office they can use polls to keep in touch with opinion trends in their constituencies. As of 1953, 48 members of the U.S. House of Representatives were known to be sponsoring opinion surveys, and 14 more said they intended to do so; significantly, most lawmakers in question were in the younger age groups, a fact that suggests that more congressmen may be using polls at the present time (Hawver 1954, p. 125). Politicians and political parties in western European countries are known to have shown similar interests in survey research.

Private businesses and associations have made even more frequent use of polling techniques. Most of the surveys done for business come under the heading of market rather than opinion research and are primarily concerned with product preferences, but large numbers are concerned with public issues, such as government regulation, the economic outlook, or community relations. There are also numerous studies of the public image of individual enterprises or whole industries. Such studies fall somewhere between market and opinion research, and many of them are commissioned by public-relations counsel working on behalf of the business or industry concerned. Labor unions, churches, and professional bodies have also used polls on occasion; the medical profession has been especially active in this regard. Innumerable small polls (and some large ones) have been conducted by academic researchers working with students. Indeed, it has been said that more is known about the opinions of American college sophomores than about the opinions of any other group in the world.

A very different but equally important use of measurement is made by those who are primarily interested in other approaches to public-opinion research. Quantitative studies have provided many of the building blocks for those who study the internal relationships among opinions on public issues, the political role of public opinion, and the impact of communications. Measurement techniques have thus contributed importantly to the formation of theories and hypotheses in all branches of public-opinion research.

### Public opinion as a form of organization

Long before techniques for systematically measuring opinions were developed, it was noted that public opinion seemed to have qualities that made it something more than the sum of individual opinions on an issue. It was presumed to have a

force and vitality unconnected with any specific individual. The German poet Wieland (1799) spoke of it as "an opinion that without being noticed takes possession of most heads," and the sociologist Ferdinand Tönnies ([1887] 1957, p. 256) observed that "whatever may come to be considered a public opinion, it confronts the individual with an opinion which is in part an extraneous power." Some scholars postulated the existence of a "group mind" with a will of its own, and observation of crowd behavior seemed to confirm the existence of some psychic entity that could seize hold of many individuals at once and lead them to behave in ways that no one of them would have behaved under other circumstances [see COLLECTIVE BEHAVIOR].

The concept of a group mind was soon discarded, since no empirical referent for it could be found, but the search continued for an explanation of the way public opinion differed from a simple summation of individual opinions. One explanation that has been advanced by a number of twentieth-century social scientists is that individual opinions are sometimes related to each other in such a way as to form a kind of organization. Charles Horton Cooley described public opinion as "no mere aggregate of separate individual judgments, but an organization, a cooperative product of communication and reciprocal influence" ([1909] 1956, p. 121). Ideas such as these have resulted in the abandonment of the search for an entity or content labeled "public opinion" that can be discovered and then analyzed; emphasis has been placed instead on the study of multi-individual situations and of the relationships among the opinions held by various people in these situations (Allport 1937, p. 23).

If public opinion is viewed as a species of organization or as a bundle of relationships, questions arise as to what the nature of these relationships is, how they are formed, how they persist, and why they dissolve. The relationship most frequently examined is that between leaders and followers—that is, between political influentials and the mass of the people. A number of writers have analyzed the techniques by which politicians and statesmen are able to develop a common will among masses of disparate individuals by manipulating concepts and symbols (Lippmann 1922; Lasswell 1927; 1935). Woodrow Wilson, for instance, used his Fourteen Points to mobilize a common opinion out of the wide varieties of opinions churned up by World War I. Later investigations have found that opinion leaders, or those who are influential in determining what others think about current issues,



are not concentrated only at the top of the social and political pyramids but can be found throughout the population. Each socioeconomic group has its own opinion leaders, who play an important part in determining the attitudes of other members of their group. The same leaders are not necessarily influential in all subject areas, however; one may be considered an authority on political questions, another on economic questions, and so on (Katz & Lazarsfeld 1955). Studies of the way in which group membership influences individual opinions have contributed substantially to understanding the relationships involved in public opinion. For example, researchers have found that soldiers transferred from one unit to another during World War II adopted attitudes prevalent in the unit to which they had been transferred (M. B. Smith 1949).

Knowledge about the internal structure of public opinions, nevertheless, is still limited and lags far behind measurement. This is largely because of the difficulties involved in this type of research, both in conducting experiments that enable the investigator to make systematic observations of the relationships among those holding individual opinions and in quantifying such observations as can be made about these relationships in real-life situations, outside the laboratory. Despite the relatively undeveloped state of our knowledge about the internal relationships among opinions, some of the implications of this approach have found recognition among practitioners engaged in trying to influence popular thinking. It is common for public-relations specialists or propagandists to compile lists of "influentials," often on the basis of sociological criteria, on the assumption that ideas reaching people on these lists will spread to a wider public. Some public-relations practitioners have spoken of their task as the "crystallization" of public opinion—that is, the transformation of individual attitudes into a collectivity that can exert influence (Bernays 1923).

People concerned with building viable democratic polities in new nations may also find it useful to think of public opinion as a form of organization. Students of developing areas have noted that private citizens who take an interest in political questions in emerging nations are frequently out of touch with each other and unable to interact constructively (Shils 1963). There is thus no infrastructure of private organizations and public opinion between the government and the population masses, and this lack tends to facilitate sudden and radical shifts in governments and policies. These shifts, it is hypothesized, would be less extreme if a way could be found to engage all those with po-

litical interests, both in and out of government, in a dialogue that would lead to an increasing degree of consensus on important national issues. The problem is thus to relate individual opinions to each other in such a way that fairly stable bodies of opinion, capable of exerting political influence on each other and on the government, will be formed.

### The political role of public opinion

Political scientists and historians have been most interested in the part that public opinion has played and is playing in political life. Accordingly, they have looked upon it primarily as an expression of opinion from the public that reaches the government and that the government finds prudent to heed (Speier 1950; Key 1961, p. 14). Some students have added the concept of latent public opinion, that is, public opinion that governmental officials expect will form if they do or do not do something and that thus influences their actions even though it has not yet taken shape (Truman 1951, pp. 511–512). This approach may bypass such questions as how opinions are distributed and what kinds of relationships they have with each other, but it does not necessarily do so. Numerous political thinkers have been interested in the measurement and organizational structure of public opinion.

Although the term "public opinion" was not used prior to the eighteenth century, historians have identified phenomena very much like it in both ancient and medieval civilizations (Bauer 1929), and the relationship between government and mass opinion receives attention in the work of Plato, Aristotle, and other classical as well as medieval writers. Following the Protestant Reformation and the Renaissance in Europe, both of which resulted in more widespread and intensive discussion of competing beliefs and ideas, popular opinion was increasingly seen as playing a part in governmental decisions. Machiavelli believed that princes should take this opinion into account as one element in their calculations, and by the seventeenth and eighteenth centuries leading political philosophers were paying tribute to its power. Rousseau held that all laws were ultimately based on public opinion and regarded the free expression of it as a major safeguard against despotism. Bentham stressed that the legislator could not ignore it (Palmer 1936).

By the nineteenth century the concept of public opinion had entered into the mainstream of political theory. In Europe it was frequently seen as a weapon of the middle class against the old order.

Friedrich J. Stahl, a Prussian conservative, described public opinion as the "will of the middle class" and believed that with the aid of the press the middle class would prevail over the crown. Johann Kaspar Bluntschli, the Swiss legal scholar, wrote that public opinion was "predominantly the opinion of the middle classes, which form their own judgments and express these in unison" (Lenz 1956, p. 60). James Bryce saw public opinion in western Europe as being substantially the opinion of the class that "wears black coats and lives in good houses" (1888, p. 260 in 1891 edition). Observers of the United States held that public opinion there rested on a far wider population base. Bryce, writing in the 1880s, described the American system as government by public opinion and believed that popular attitudes were expressed primarily by the press, political parties, and elections. Half a century earlier Tocqueville had likewise considered the mass basis of American public opinion, and he saw in its influence a threat to independence of thought: "In America, the majority raises formidable barriers around the liberty of opinion" (1835, p. 117 in 1956 edition); "I am not the more disposed to pass beneath the yoke because it is held out to me by the arms of a million of men" (*ibid.*, p. 149).

The current concerns of those who are primarily interested in the interplay between public opinion and government center on questions defined in the previous century, but the discussion of these questions customarily draws upon quantitative data from opinion polls, content analyses, or other sources. A large portion of this literature consists of arranging and analyzing poll data in order to disclose the content and characteristics of opinions with which governments are or should be concerned (Key 1961). Thus, one can find a substantial number of studies that deal with public opinion and foreign policy, public opinion and social legislation, or public opinion and economic issues. Many of these studies focus on population subgroups of presumed political significance.

Closely related to analyses of public opinion on political issues are studies that are concerned with the ways in which public opinion influences government. In addition to using poll data, these studies usually describe the activities of pressure groups and political parties and the techniques of individuals who attempt to affect government policy (for a selective bibliography, see Childs 1965, pp. 257-260). Included in this group are analyses of mail reaching executive or legislative officials. Although most such treatments have dealt with the United States, there is a growing body of

literature both on pressure groups in other countries and on efforts by individuals in other countries to participate in official decisions. A recent study compared citizens' attitudes toward their own ability to influence governmental policy in West Germany, Italy, Mexico, and the United States (Almond & Verba 1963).

The weakness of most studies dealing with the influence of public opinion on government policy is that they cannot trace the effectiveness of the attempts that are made to exert influence. The equation of activity with effectiveness is simply assumed. Nevertheless, there are some notable exceptions. For example, the West Coast fishing industry was found to play a major part in the negotiation of a settlement with Japan following World War II (Cohen 1957).

The other side of the coin—government efforts to influence public opinion—has also been dealt with extensively. Tools most frequently used by governments for this purpose include publicity, propaganda, censorship, and a number of techniques of news management (Childs 1965, pp. 305-308). Although the manipulative attempts of government in the United States have received the lion's share of the attention of scholars, the rulers of totalitarian states have been most active in trying to mold public opinion. Efforts of the Nazi government to control public attitudes in Germany were notorious, and Soviet domestic propaganda has also received fairly extensive attention (Inkeles 1950). More recently, scholars have begun to focus on the activities of Communist China in this field (Yu 1964). As with studies of popular efforts to influence governments, treatments of governmental efforts to manipulate opinion have ordinarily been unable to show a clear connection between cause and effect. A possible exception is the case of the Soviet Union, where even anticommunist citizens appear to have adopted many of the principal categories of thought prescribed for them by the government.

Cutting across all these aspects of the relationship between government and public opinion are studies of voting behavior. These have registered the distribution of opinions on a wide variety of issues, have explored the impact of special interest groups on election outcomes, and have contributed to our knowledge about the effects of government propaganda and policy. Since voting is the method by which the largest numbers of citizens of any democracy participate in policy, analysis of the behavior of potential voters during and between election campaigns contributes fundamentally to an understanding of the democratic process. Those



who believe in democracy often find the results of such studies discouraging, in that they may show widespread prejudice and ignorance or little popular appreciation of important issues, but the results are rarely irrelevant.

To generalize about the effects of the enormous amount of theorizing, philosophizing, and research that has been done on the relationship between public opinion and government is a difficult task. It seems safe to assume that the behavior of some political leaders has been influenced by political theories that are based in part on thinking about public opinion. It can also be argued that public-opinion research has enhanced the prospects for democratic government by helping to acquaint major groups in the population with each other's attitudes and values. When, however, one looks for *specific* examples of ways in which the study of public opinion has affected the political process, the examples one finds are likely to be relatively minor and to involve measurement rather than theory. For instance, officials have sometimes been able to discount the importance of what they hear from pressure groups when they have learned about the distribution of the opinions of the population as a whole on the same issues. It is also probable that both governmental manipulators of public opinion and private groups seeking to influence official policy have been able to conduct their activities with greater sophistication because of increased knowledge about the nature of the relationship between the government and the public.

### Communication research

The task of communication research has been defined by Lasswell as that of answering the question: who says what to whom through what channel and with what effect? ([1948] 1964, p. 37). Thus, communication researchers who follow this formula have not focused their attention directly on opinions regarding public issues, but they have made a number of important indirect contributions to the understanding of public opinion. In particular, studies of symbol manipulators (those who speak), audiences (those who are exposed to communications), the role of the mass media, and the effects of the ideas that are communicated are relevant to such questions as how opinions take hold among large numbers of people, why they are distributed as they are, how they are related to each other, and how they change.

Those who are concerned with the ways in which ideas are spread have studied the activities of government spokesmen, private publicists, and propaganda and news organizations as well as the rela-

tionships between communications and policy. Few adequate generalizations have been made about propagandists. One analysis found them to have abnormally strong cravings for power, highly extroverted personalities, unreasoning intolerance for rivals, and a gift for organization and management when these are related to self-aggrandizement (B. L. Smith et al. 1946). This unflattering characterization was, however, based in large part on attention to Nazi, Fascist, and Communist spokesmen. Modern propaganda organizations are characterized by elaborate mechanisms for setting policy, operating media, studying audiences, and evaluating effectiveness (Davison 1965). The task of coordinating the output of such vast organizations with the policy of the sponsoring group (usually a government) has frequently led to clashes between propagandists and policy makers.

Studies of news media have indicated that they, too, tend to structure opinions, even though usually unintentionally. One way this is done is by giving large numbers of people a common focus of attention. An operating definition of news used by most journalists is that a subject is newsworthy if it is already in the headlines. Thus, there tends to be a circular reinforcing process, in that a subject is featured if it is already being given attention and it is given more attention because it is featured (Cohen 1963). News media also help to assure cohesion among members of major population groups. In many nations there is a "prestige paper" that is read by a large proportion of influential persons and that keeps members of this group of influentials informed about what others in the group are thinking. It also serves as a forum for exchange of ideas among them. In 1961, for example, a survey found that the *New York Times* was subscribed to by 60 per cent of American news editors, 46 per cent of utility executives, 30 per cent of college presidents, and 28 per cent of bank officers throughout the country (Kraft 1961). Newspapers such as *Le Monde*, *Pravda*, and *The Times* of London may occupy an even more central position.

Propagandists and the mass media do not, however, have it all their own way. Audience studies have shown that people give their attention selectively to the communications that are available to them and that they frequently derive meanings from these that are quite different from the ones intended. Even very intensive publicity drives have failed to increase knowledge about given subjects when people have no interest in acquiring more information. This was found to be the case in Cincinnati shortly after World War II, when civic

organizations and mass media experimented with a campaign to stimulate interest in the UN; in spite of the fact that the entire population was deluged with information about the world body, those who were uninformed prior to the campaign remained uninformed (Star & Hughes 1950). Conversely, news of events that are not given extensive publicity can circulate with amazing rapidity by word of mouth. Attention seems to be governed primarily by two factors: habit, in that people grow accustomed to relying on certain sources for information, and interest, in that people tend to look for information that they think will help them satisfy their wants or needs [see COMMUNICATION, MASS, *article on EFFECTS*].

As far as the effects of communications on opinions are concerned, a major finding of researchers has been that well-formed attitudes are highly resistant to change. Even the most skillful propaganda or advertising is unlikely to bring about shifts in attitudes that are firmly based on a person's own experience or that are shared by those with whom he comes in frequent and close contact. On the other hand, casual attitudes can be changed fairly easily. In nonpartisan local elections, where the candidates are not well known, even a small amount of information or an endorsement by a respected newspaper can sway a substantial number of votes. Similarly, the editorial stand of a newspaper can sometimes tip the scales for or against an issue in a community where opinions are fairly evenly divided and feelings do not run high. Communications can also reinforce existing attitudes and activate latent ones. Agitators frequently make use of activating and reinforcing communications in order to whip up enthusiasm for an idea and bring people to the point where they are willing to vote, demonstrate, or take some other action. Such effects may also be achieved unintentionally. During 1955, when a number of West German newspapers started to publicize crimes committed by American military personnel in West Germany, largely as a circulation-building device, they whipped up latent emotions to a point where at least one town council requested the withdrawal of American forces from the area. It later turned out that the "crime wave" represented no more than the usual number of violations with which American and West German authorities were customarily plagued and that the press had not intended to stimulate political action.

Insights from communication research are used in both the political and the commercial realms. Political uses are made most frequently by those concerned with propaganda and psychological war-

fare. In democratic elections, for instance, the political propagandist frequently tries to find an issue that is favorable to his side or unfavorable to the opposition and about which there are widespread latent attitudes; he then tries to activate these attitudes by means of emphasizing the issue in the public media. The psychological-warfare specialist often seeks the "marginal man," one who is in the opponent's camp without being firmly dedicated to it, and makes him the principal target for propaganda. In the commercial world, communication research makes it possible for the advertiser to select media that are already being given attention by the audience he wishes to reach and also to test the effectiveness of varying advertising appeals. The same principles of selection and testing can be used by those who are merchandising political ideas.

One of the most important applications of communication research is in connection with national development. The mass media play a critical role in turning traditional societies into modern ones by preparing people psychologically for life in an industrial society, by helping to build new political institutions engaging in mass education, and by providing information that is necessary for economic development (Lerner 1958; Conference on Communication . . . 1963). Accordingly, modernizing nations have been advised to examine the channels for the flow of information within their borders and to plan for the growth of the mass media in phase with the growth in other sectors. The mass media are necessary for development not only because they link people with government and disseminate information that is needed in educational and economic programs but also because they draw diverse people together around common national problems and interests and make it possible for them to participate in public affairs (Schramm 1964).

### Future prospects

Although there is little agreement on a precise definition of "public opinion," and many of its facets remain hazy, constantly increasing use of the term during the past two centuries testifies to its utility. As societies have become more complicated and ever larger masses of people have become involved in political and economic life, students have felt a growing need for some concept that refers to collectivities falling between the undifferentiated mass, on the one hand, and primary groups and formal organizations, on the other. In modern society people do not behave like isolated individuals, but neither can their behavior



be explained exclusively in terms of their organizational and group membership. This became abundantly apparent just prior to and during the French Revolution, when the middle class began acting much as though it were an organized force, even though no principle of organization was immediately apparent. True, the middle class had its salons and coffeehouses, where information and ideas were exchanged, and sometimes its newspapers, but were these communication facilities alone sufficient to threaten the stability of the *ancien régime*? The ferment of the times could be explained only in terms of some greater force, and this force was labeled public opinion.

As a result of increases in literacy and proliferation of communication media, public opinion soon ceased to be exclusively a middle-class phenomenon in the principal industrialized areas of the world. Less-advantaged members of the community have also been able to achieve a consensus on some issues, although less frequently and intensely than those at higher socioeconomic levels. The civil rights movement in the United States has succeeded in mobilizing public opinion far beyond the middle class. It is probable that public opinion in the developing countries will evolve in similar fashion but will take less time to do so.

Public opinion has played a smaller role in communist countries than in the democracies. The efforts of the state to control the principal channels of public communication and to prevent nonofficial links among the citizenry have made it more difficult for individual opinions to become related and for consensuses to develop. Nevertheless, this has occurred in some instances, either because independent-minded artists and writers have given expression to widely shared ideas or because it has been made possible by word-of-mouth communication. Both rebellious writers and interpersonal channels seem to have played a role in the growth of public opinion that led to the antiregime uprising in Hungary in 1956. Communist spokesmen occasionally assert that public opinion does exist in countries having a totalitarian-socialist form of government but that it is a common opinion in which the entire population shares and is adequately expressed in single-list elections and mass demonstrations. It is improbable that such asseverations are made seriously, especially since public-opinion studies that recognize important differences among individuals and population groups have been reported from eastern Europe and the Soviet Union with increasing frequency.

Our understanding of the role that public opinion plays in political, economic, cultural, and other

areas can be expected to increase during the coming years as research and thinking at all levels is pursued throughout the world. Further comparisons of public-opinion processes in various societies would be especially helpful. Even more important would be empirical and theoretical studies linking together the four approaches described above, so that each would contribute to an interlocking and mutually reinforcing whole. Only when this integration has been achieved will a coherent theory of public opinion be possible.

W. PHILLIPS DAVISON

[See also ATTITUDES; CENSUS; COMMUNICATION, MASS; COMMUNICATION, POLITICAL; DIFFUSION, article on INTERPERSONAL INFLUENCE; MASS SOCIETY; PROPAGANDA; PUBLIC RELATIONS; REFERENCE GROUPS; and the biographies of BRYCE; COOLEY; DICEY; MACHIAVELLI; TOCQUEVILLE.]

#### BIBLIOGRAPHY

- ALBIG, WILLIAM 1956 *Modern Public Opinion*. New York: McGraw-Hill.
- ALLPORT, FLOYD H. 1937 Toward a Science of Public Opinion. *Public Opinion Quarterly* 1:7-23.
- ALMOND, GABRIEL A. (1950) 1960 *The American People and Foreign Policy*. New York: Praeger.
- ALMOND, GABRIEL A.; and VERBA, SIDNEY 1963 *The Civic Culture: Political Attitudes and Democracy in Five Nations*. Princeton Univ. Press.
- BAUER, WILHELM 1929 *Die öffentliche Meinung in der Weltgeschichte*. Potsdam (Germany): Akademische Verlagsgesellschaft Athenaion.
- BERNAYS, EDWARD L. 1923 *Crystallizing Public Opinion*. New York: Boni & Liveright.
- BRYCE, JAMES (1888) 1931-1933 *The American Commonwealth*. 3d ed. 2 vols. New York and London: Macmillan. → An abridged edition was published in 1959 by Putnam.
- BUCHANAN, WILLIAM; and CANTRIL, HADLEY 1953 *How Nations See Each Other: A Study in Public Opinion*. Urbana: Univ. of Illinois Press.
- BUREAU OF SOCIAL SCIENCE RESEARCH, WASHINGTON, D.C. 1956 *International Communication and Political Opinion: A Guide to the Literature*, by Bruce L. Smith and Chitra M. Smith. Princeton Univ. Press.
- CANTRIL, HADLEY 1966 *The Pattern of Human Concerns*. New Brunswick, N.J.: Rutgers Univ. Press.
- CHILDS, HARWOOD L. 1965 *Public Opinion: Nature, Formation and Role*. Princeton, N.J.: Van Nostrand.
- COHEN, BERNARD C. 1957 *The Political Process and Foreign Policy: The Making of the Japanese Peace Settlement*. Princeton Univ. Press.
- COHEN, BERNARD C. 1963 *The Press and Foreign Policy*. Princeton Univ. Press.
- CONFERENCE ON COMMUNICATION AND POLITICAL DEVELOPMENT, DOBBS FERRY, N.Y., 1961 1963 *Communications and Political Development*. Edited by Lucian W. Pye. Princeton Univ. Press.
- COOLEY, CHARLES H. (1909) 1956 *Social Organization: A Study of the Larger Mind*. In Charles H. Cooley, *Two Major Works: Social Organization and Human Nature and the Social Order*. Glencoe, Ill.: Free Press. → Each title reprinted with individual title page and

- pagination. Separate paperback editions were published in 1962 by Schocken.
- DAVISON, W. PHILLIPS 1965 *International Political Communication*. Published for the Council on Foreign Relations. New York: Praeger.
- ERSKINE, HAZEL G. 1963 *The Polls: Exposure to International Information*. *Public Opinion Quarterly* 27: 658-662.
- FREE, LLOYD A. 1967 *The Role of Public Opinion in International Relations*. Unpublished manuscript.
- GALTUNG, JOHAN 1964 Foreign Policy Opinion as a Function of Social Position. *Journal of Peace Research* 1:206-231.
- HAWVER, CARL 1954 *The Congressman and His Public Opinion Poll*. *Public Opinion Quarterly* 18:123-129.
- HYMAN, HERBERT H.; and SHEATSLEY, PAUL B. 1947 Some Reasons Why Information Campaigns Fail. *Public Opinion Quarterly* 11:412-423.
- HYMAN, HERBERT H.; and SHEATSLEY, PAUL B. 1964 Attitudes Toward Desegregation. *Scientific American* 211, July: 16-23.
- INKELES, ALEX (1950) 1958 *Public Opinion in Soviet Russia: A Study in Mass Persuasion*. 3d printing, enl. Russian Research Center Studies, No. 1. Cambridge, Mass.: Harvard Univ. Press.
- INSTITUTE FOR INTERNATIONAL SOCIAL RESEARCH 1961 *Some International Implications of the Political Psychology of Brazilians*. Princeton, N.J.: The Institute.
- KATZ, ELIHU; and LAZARSFELD, PAUL F. 1955 *Personal Influence: The Part Played by People in the Flow of Mass Communications*. Glencoe, Ill.: Free Press.
- KEY, V. O. JR. 1961 *Public Opinion and American Democracy*. New York: Knopf.
- KLAPPER, JOSEPH T. 1960 *The Effects of Mass Communication*. Glencoe, Ill.: Free Press.
- KRAFT, JOSEPH 1961 *The Future of the New York Times*. *Esquire* 55, no. 4:121-122.
- LASSWELL, HAROLD D. (1927) 1938 *Propaganda Technique in the World War*. New York: Smith.
- LASSWELL, HAROLD D. 1935 *World Politics and Personal Insecurity*. New York and London: McGraw-Hill. → A paperback edition was published in 1965 by the Free Press.
- LASSWELL, HAROLD D. (1948) 1964 *The Structure and Function of Communication in Society*. Pages 37-51 in Institute for Religious and Social Studies, Jewish Theological Seminary of America, *The Communication of Ideas: A Series of Addresses*. Edited by Lyman Bryson. New York: Cooper Square.
- LAZARSFELD, PAUL F.; BERELSON, BERNARD; and GAUDET, HAZEL (1944) 1960 *The People's Choice: How the Voter Makes Up His Mind in a Presidential Campaign*. 2d ed. New York: Columbia Univ. Press.
- LENIN, V. I. 1929 *Agitation und Propaganda*. Berlin: Verlag für Literatur und Politik. → Published posthumously. Contains a collection of Lenin's earlier writings on the subject first published in Russian.
- LENZ, FRIEDRICH 1956 *Werden und Wesen der öffentlichen Meinung*. Munich: Pohl.
- LERNER, DANIEL 1958 *The Passing of Traditional Society: Modernizing the Middle East*. New York: Free Press. → A paperback edition was published in 1964.
- LIPPMANN, WALTER (1922) 1944 *Public Opinion*. New York: Macmillan. → A paperback edition was published in 1965 by the Free Press.
- MALETZKE, GERHARD 1963 *Psychologie der Massenkommunikation: Theorie und Systematik*. Hamburg (Germany): Verlag Hans Bredow-Institut.
- MICHIGAN, UNIVERSITY OF, SURVEY RESEARCH CENTER 1964 *The American Public's View of U.S. Policy Toward China: A Report*. New York: Council on Foreign Relations.
- PALMER, PAUL A. 1936 *The Concept of Public Opinion in Political Theory*. Pages 230-257 in *Essays in History and Political Theory in Honor of Charles Howard McIlwain*. Cambridge, Mass.: Harvard Univ. Press.
- PERRY, PAUL 1960 Election Survey Procedures of the Gallup Poll. *Public Opinion Quarterly* 24:531-542.
- SCHRAMM, WILBUR L. 1964 *Mass Media and National Development: The Role of Information in the Developing Countries*. Stanford Univ. Press.
- SHILS, EDWARD 1963 *Demagogues and Cadres in the Political Development of the New States*. Pages 64-77 in Conference on Communication and Political Development, Dobbs Ferry, N.Y., 1961, *Communications and Political Development*. Edited by Lucian W. Pye. Princeton Univ. Press.
- SMITH, BRUCE L.; LASSWELL, HAROLD D.; and CASEY, RALPH D. 1946 *Propaganda, Communication and Public Opinion: A Comprehensive Reference Guide*. Princeton Univ. Press.
- SMITH, M. BREWSTER 1949 *The Combat Replacement*. Pages 242-289 in Samuel A. Stouffer et al., *The American Soldier*. Volume 2: Combat and Its Aftermath. Princeton Univ. Press.
- SPEIER, HANS (1950) 1952 *The Historical Development of Public Opinion*. Pages 323-338 in Hans Speier, *Social Order and the Risks of War: Papers in Political Sociology*. New York: Stewart. → First published in Volume 55 of the *American Journal of Sociology*.
- STAR, SHIRLEY A.; and HUGHES, HELEN MACGILL 1950 Report on an Educational Campaign: The Cincinnati Plan for the United Nations. *American Journal of Sociology* 55:389-400.
- STOUFFER, SAMUEL A. (1955) 1963 *Communism, Conformity, and Civil Liberties: A Cross-section of the Nation Speaks Its Mind*. Gloucester, Mass.: Smith.
- TOCQUEVILLE, ALEXIS DE (1835) 1945 *Democracy in America*. 2 vols. New York: Knopf. → First published in French. Paperback editions were published in 1961 by Vintage and by Schocken.
- TÖNNIES, FERDINAND (1887) 1957 *Community and Society (Gemeinschaft und Gesellschaft)*. Translated and edited by Charles P. Loomis. East Lansing: Michigan State Univ. Press. → First published in German. A paperback edition was published in 1963 by Harper.
- TRUMAN, DAVID B. (1951) 1962 *The Governmental Process: Political Interests and Public Opinion*. New York: Knopf.
- WARNER, LUCIEN 1939 *The Reliability of Public Opinion Surveys*. *Public Opinion Quarterly* 3:376-390.
- WIELAND, CHRISTOPH MARTIN 1799 *Gespräch unter vier Augen*. Leipzig: Göschen.
- YU, TE-CHI 1964 *Mass Persuasion in Communist China*. New York: Praeger.

## II

## POLITICAL OPINION

"Political opinion" is a convenient term for referring to the context of authoritative decision, governmental and nongovernmental, in which persons in positions of discretionary authority anticipate or respond to estimates of the community's demands, expectations, or requirements. As the sam-



ple survey and other instruments of social science research add precision to the concept of public opinion—defined as the distribution of personal opinions on public objects in the population outside the government—it becomes increasingly apparent that the opinion process is not identical with or completely determinative of the judgments of persons in a position to formulate and shape public policy. The concept of political opinion is thus a shorthand expression for the relation or sets of relations between the opinion-forming and policy-making processes in society. In empirical terms, it leads to the following question: which opinions and views, held by which persons outside the government, are heeded by which persons who make authoritative policy decisions? Philosophically, it also raises the question of meaning: what function does public opinion play in the political process? Public opinion is sometimes considered to be a sanction (legitimizing symbol), sometimes an instrument (data), and sometimes a generative force (directive and limit) in the policy process. Clearly, then, normative assertions about “government by public opinion” do not accurately describe the processes by which opinions are actually utilized by the makers of public policy.

### Opinion in the public context

In order to understand the many ways in which public opinion actually functions in the political process, it is necessary to clarify the meaning of “opinion” and “public.” “Opinion” includes the whole range of personal and private sentiments, beliefs, and systematic views, felt and expressed, rational and nonrational, that both affect the legitimacy and stability of the social order and political system and condition the feasibility and acceptability of public policies. In a political setting, the relevant criterion is not the truth or falsity or the rightness or wrongness of opinions but their significance for the electoral success or failure of officeholders and candidates, for the degree of popular support for existing and proposed policies aimed at the security and prosperity of the community, and ultimately for the viability of the constitutional and political system.

Insofar as opinions are expressed, it is useful to distinguish two categories of opinion statements: (1) those of *preference*, which include expressions of individual feeling, conviction, and value; and (2) those of *fact*, which purport to describe transpersonal reality in objective terms of verifiable evidence. The important characteristic of opinion statements, as distinct from postulates of knowledge, mystical or religious faith, and matters of

belief beyond discussion (e.g., incest), is that they are recognized as expressing controversial, conflicting ideas, capable of being accepted by rational minds, and hence are appropriate for discussion among individuals and groups. Discussion can lead to agreement and can settle differences of opinion; only violence, authority, or an appeal to a metaphysical or a divine source can resolve differences in objects of faith and belief, which are literally nondebatably and thus fall outside the sphere of opinion.

There is a notable tradition in the history of normative thought about the relation of the public and the private in matters of opinion. Milton's *Areopagitica*, Locke's *Letter Concerning Toleration*, and Mill's *On Liberty* are classic statements of the moral and political problems posed by the personal freedoms of religious belief, thought, and expression; by a press that is not owned or controlled by the government; and by the respective rights of the majority and of minorities.

Through most of the nineteenth century the dominant tradition assumed that the public sphere was defined by the legally coercive acts of government officials acting in the name of the monarch or representing the people organized as a political community. The work of Marx, James, and Freud, coupled with the rise of the “culture” concept in anthropology, shifted the focus of attention to the economic and social-psychological determinants of opinion, almost to the point of obliterating the distinction between “public” and “private.” Some writers treated the public and political aspects of opinion as logical derivatives of class or status ideologies; others, as by-products of the personality-forming processes by which individuals are adjusted to society. The term “public” was given the sense of “interpersonal relations.” For example, Charles Cooley saw public opinion as the product of social communication and reciprocal influence; John Dewey, as the consequences of interpersonal acts upon persons other than those directly involved in transactions. But these concepts turned out to be virtually identical with the notion “group,” measured by “the extent and complexity of interpersonal relations among the members of specified aggregates, . . . which under specified conditions exhibit organization of varying kinds and degrees” (Lasswell & Kaplan 1950). Hence “public” had to be differentiated by specifying the requirement of a more inclusive consensus, within which various opinion aggregates could coexist.

These theoretical developments emphasized the roots of public opinion in private and group experience and demonstrated the futility of absolute

differentiations between "individual," "group," and "public." However, the conception of the public as a simple extension of interpersonal contact and communication remains unsatisfactory. First, it is not clear that an opinion consensus is a necessary or sufficient condition for a public to exist. Second, the problem of coercion is ignored. Finally, the linkages and adjustments connecting group opinions with the formulation of public policy are vaguely referred to under the heading "pressures" or "compromise." Models of the political process that explicitly recognize the ambiguous relation between conflicting opinion preferences and actual policy decisions are still required.

### Theories of political opinion

The range of individual and group opinions taken into account by political rulers, representatives, and public officials varies widely among political societies, historical and contemporary. In primitive communities and in feudal, caste, or closed-class social systems, the range of political communication seems to have been narrowly confined within limits set by ritual, custom, and social structure. In fractionalized, transitional cultures the ruling group often restricts sharply the scope and subjects of public discussion and indeed controls more or less completely the channels of mass communication. Since World War I it has become fashionable for all governments to claim to be democratic, in the sense that a high degree of correspondence is imputed to exist between the acts of government officials and the majority opinion. However, we may draw a preliminary, fundamental distinction between "constitutional" and "people's" democracies, in terms of the degree of non-governmental control of the instruments of opinion formation and the relation that is asserted to exist between public opinion and governmental decision making.

Under constitutional governments, public officials operate through prescribed, legal procedures for issuing and validating statements of public policy. Nongovernmental institutions of public information and discussion are allowed ample opportunity, before, during, and after the official process, to formulate the issues and to develop a patterned distribution of individual responses from "the country." All this goes on in an atmosphere free from military or private violence, and individuals and groups are free to criticize and respond to governmental programs of information and persuasion. In this context the acts of public officials are interpreted as symbolic, authoritative expressions of the public interest, legally enforceable

until amended or rescinded by proper constitutional procedures after a period of public reaction, political agitation, and response.

Contrariwise, in so-called people's democracies it is held that the will of the "people," the "masses," the "nation," or the *Volks* is known or available only to the ruler or the party; this "will" is embodied in the ruler's or the party's acts, and only such information and views as are found necessary or appropriate are legitimate matters for public knowledge and discussion. To this end, the channels of public communication and conditions of group association are rigorously limited and controlled by governmental and party officials.

Speculative theories of political opinion can be grouped into five classes; each has affinities to one of the two basic types above but may also resemble the other in some features.

**Elitist theories.** Perhaps the oldest theory of political opinion is that of the elite, or ruling class, according to which society is horizontally divided into leaders and followers and vertically divided into hierarchies of racial, functional, socioeconomic, or demographic groups. The ruling class is composed of the leaders of the several vertical-group hierarchies. The consensual foundations of political community, according to elitist theorists, lie in ethnic ties, qualities of the people, ancient customs, glorious historical events, and symbolic documents embodying self-limiting agreements between the nominal ruler and the elite leaders. In the elitist conception of political opinion, the status and opinion leaders outside government are seen as occupying a hierarchical position above the points of official decision, so that the lower-status governmental officials, by definition, simply execute the decisions made by the ruling class. Representative of this approach is the work of Gaetano Mosca [see MOSCA].

**Absolute-sovereignty theories.** A second class of theories of political opinion includes those absolutist doctrines of sovereignty, both monarchic and popular, that view society as an abstract mass of individuals. In the absence of rulers, the people are able to act only chaotically or in compulsive unanimity, for example, as a mob. The abstract "people," therefore, find it necessary to abdicate power to a no less abstract absolute ruler, monarch, leader, or party. These entities alone are capable of knowing and expressing the mystical, unifying will variously named *volonté générale*, *Volksgeist*, sovereign people, and mass proletariat. Interpreters of Hobbes, Rousseau, and Hegel can point to paragraphs and sections showing that these thinkers were aware that the social structure and opinion



processes of society are more complex than the simple dichotomy of people and ruler; nonetheless, in answer to the question of where the source of supreme power in the political community is located, each postulated an unlimited governing authority, either in the mass-majoritarian consensus of the people or in the moral and political responsibilities of the state-ruler. None of the Leninist-Stalinist theories (or their adaptations by Mao Tse-tung) about the relations of the party to the masses overcomes the difficulty of the people-state dilemma. [See SOVEREIGNTY.]

**Manipulative theories.** Closely associated historically with the concept of popular sovereignty is the idea that people have to be educated, guided, induced, or persuaded. In practical terms, this has led to the notion that issues can be created and that new cleavages and changes in the mass distribution of opinion can be brought about by events and adroit acts of opinion leaders, politicians, prestige figures, or persons who control the content of the mass media. War propaganda during World War I, the tremendous increase in commercial advertising and public relations activities thereafter, the deliberate use of propaganda and agitation by totalitarian and revolutionary mass movements, and the conscious development of psychological warfare in World War II, all highlighted what may be called the mass-manipulation theory of political opinion. During the interwar period there was a tendency to overstate the degree to which modern "mass society" can be controlled without legal and political monopolies over the press and other mass media. However, there was also a clear realization that social forces and public opinion are not independent, exogenous variables in the opinion-policy process but are affected by, and indeed are sometimes produced by, leaders and events. [See, for example, LIPPMANN.]

**Constitutional-democratic theories.** The fourth category of political opinion models contains those theories of representative democracy which postulate the idea of a covenant or agreement among people who, by giving up their prepolitical freedom of action, are enabled to establish a government of limited powers in order to secure certain values (union and independence, security of life and property, justice and liberty, the common welfare) that individuals without government cannot obtain by themselves. [See *The Federalist*; see also LOCKE; SOCIAL CONTRACT.] The metaphorical notion of an "agency" contract allows the imposition of substantive and procedural restraints upon the legitimacy of discretionary acts by elective representatives and

public officials. It also provides guarantees of personal and private rights, requirements for insuring conformity between governmental acts and majority opinion (free elections, freedom of speech and of the press, and the right of assembly and of petition), and a prescribed procedure for amending and revising governmental structures, procedures, and powers. Such widely varying theories of political organization as monarchy, minority or restricted majority rule, and populist democracy have been reconciled with representative-democratic forms, when they admitted prior legitimacy and loyalty to a political order based on the principles of limited powers and the consent of the governed [see LINDSAY].

**Rationalist-idealist theories.** The fifth category, the rationalist-idealist model of political opinion, was perhaps best formulated by A. L. Lowell and R. M. MacIver, who sought to identify the conditions necessary for individuals and groups to participate rationally in public affairs [see LOWELL and MACIVER]. The specified conditions were (1) that the people be politically organized and act as a community rather than as a mass or crowd; (2) that the right of minorities to hold their opinions and to organize politically for peaceful opposition be recognized but that the minority submit to and obey majority decisions; (3) that majorities and minorities alike accept and abide by the structure and processes of governmental decision, including the rules for amendment and change; (4) that the members of the political community have access to the facts reasonably necessary to arrive at a rational decision on a given issue; (5) that they engage actively in discussion and participate in public affairs, so as to be capable of assessing political realities; and (6) that they explicitly adjust their votes and other political acts to their conception of the common good and public interest. Lowell's criteria may raise doubt that any human government can be wholly democratic, but they perform the important task of clarifying the qualities of individual and group action needed if the members of a community are to act as a *public*.

**Speculative versus empirical theories.** The division between speculative and empirical theories is not precise, nor are the two mutually exclusive. Generally speaking, speculative models implicitly or explicitly incorporate the observer's preferences into his conceptual system and attempt to describe and classify human behavior and institutions directly in terms of a priori, complex categories that defy quantitative expression. Empirical or analytical models extrapolate from individual behavior a

minimum number of elemental, conceptual variables, supposedly identifiable by standard observational procedures independent of the observer's or respondent's preferences, and attempt to explain behavior indirectly by inferring (predicting or generalizing) the probable covariation between two or more such variables.

Certain obstacles confront all empirical theories of political opinion. One is the obvious methodological problem of systematically gaining access to persons in positions of public authority or private influence for purposes of observation. More important, conceptually, is the problem of variations in the setting in which the opinion-policy relation occurs. The two-way relation between opinion and policy formation can be studied on the level of, and within, the individual personality; in the interactions of two or more persons in small groups; in leader-follower behavior within large groups, whether or not they possess formal structure and bureaucratic organization; and in the complex relationships between key individuals in the multi-group process of public, official policy making. Intensive studies of the conditions under which changes can be induced in individual opinion, personality, and behavior have great potential value for research in more inclusive contexts, but the variations of scale and setting often seem to reduce the applicability of concepts employed in personality and small-group studies to analogies and metaphors on the level of political and societal behavior.

### Empirical research

On the macroscopic level of political society, studies of public opinion and policy formation do not agree about the analytical terminology for the different aspects of the opinion-policy relation, nor have they integrated the conceptual tools available. Past approaches to the study of political opinion can be grouped into four categories with respect to the focus of analytical attention: (1) the formal sanctions that justify and limit the exercise of power within the process of public opinion and policy formation; (2) the initiating sources and the environmental conditions that galvanize or inhibit the process in action; (3) the critical decision-making components in policy determination; (4) the total opinion-policy process or system, through which opinions are transmitted, policy decisions reached, and behavioral adjustments effected on the part of the members of the political society. Among the social sciences political science has been largely preoccupied with (1) and (3), soci-

ology with (2) and (4), psychology and economics with (3) and (4), and cultural anthropology with (1) and (4). Since 1945, however, interdisciplinary communication, especially in the United States, has been steadily breaking through such conceptual jurisdictions.

**Political opinion—role as sanction.** The location, organization, and exercise of formal, legal sanctions over the process of public opinion and policy formation has long been a central concern of conventional political theory, comparative government, and constitutional law and history. Not until the distinction between legal and political sovereignty was recognized, however, could public opinion become a respectable subject of academic research, and then new methods, concepts, and even disciplines became necessary.

Evidence of the role of political opinion as sanction has been found in (1) the demonstration of widely held faith among the population in the legitimacy of the structures and processes by which public policy decisions are reached; (2) the measurement of divisions of the electorate into organized parties and groups which provide support for, criticism of, or opposition to official government policies; (3) the estimates of the degree of popular satisfaction with the performance of the regime or party in power in promoting the security and prosperity of the political community; (4) the discovery of norms of behavior and procedure expected from public officials in promulgating and enforcing public policies; and (5) the analysis of the moods and qualities of opinion to which policy makers have to adjust their thinking and public behavior.

The concept of political opinion has stimulated inquiries into the distinctive calling (vocation) of the *politician*—the specialist in organizing political controversy, in formulating and discussing issues, in making decisions and taking protective policy positions in public, in return for which he obtains influence over or access to elective office. Opinion research has provided objective data on the variability of popular moods, demands, expectations; on the scope for choice, error, and misjudgment in policy making; and on career opportunities and the rewards and dangers involved in estimating opinion distributions and trends. Finally, research has documented the conception of government by public opinion as a process of identifying, investigating, and adjusting the differences between mass, interest-group, professionally political, and technically expert conceptions of the public interest, as opposed to the old-fashioned method of elaborating



the distinction between the "partial" interests of groups and the "common" interests of the public or community as a whole.

**Political opinion as articulation of interests.** Of all research into political opinion, studies of social structure and mobility, group organization and activity, and processes of interpersonal and mass communication and their effects upon opinion formation and change have made the greatest advances in precision and sophistication since 1925. However, even in these areas such research has contributed more to an understanding of the sources of public policy formation and of society's responses to the policy than to an understanding of the operation of the whole process.

Major contributions have been made toward (1) discriminating the different kinds of opinion aggregates that arouse political controversy (the mass-attention group, the reference group, the attentive public, the opinion-making public, etc.); (2) isolating and measuring the determinants of opinion by correlating self-identified group opinions with those classified by objective measures of sociopolitical stratification (multivariate analysis); (3) extending our cognitive "maps" of policy making in organizations that recruit opinion makers who articulate group interests throughout the whole field of cultural, economic, and political activity; (4) specifying the conditions under which individual and mass attitudes and behavior can be modified, distorted, or transformed by organization, leadership, and manipulation; (5) measuring the content and audience exposure of communications; and (6) analyzing the behavior of voters, consumers, buyers, and members of other functional groups in the making up and changing of their minds as the result of exposure to various forms of communication.

The sample survey has been a major vehicle of this progress, but for studying the personal transmission of opinions, the impersonal processes of organizational communication, and the relations between opinion leaders and policy makers, technicians have found it necessary to supplement the sample survey with sociometric, intensive-interview, and other observational methods. The sample survey technique has not yet enabled social scientists to construct a satisfactory general model of how social structure and group organization activate, permeate, and respond to the policy process. It has, however, enormously enhanced their ability to verify, at points of time before, during, and after the process of public decision, the extent to which information, attention, and other attributes of opinion distributions conform to hypothesized

statements, however derived. It may be conjectured that the next major advance in our knowledge of the opinion-policy relation will emerge from developments in the fields of organizational theory and decision-making behavior, where sharper concepts of interaction analysis are being developed.

**Dimension of decision makers' behavior.** Another large body of empirical inquiry into the opinion-policy process centers upon the behavior of persons in positions to make authoritative decisions that bind or control the acts of others. The study of leadership and organizational behavior falls partially into this category, as do case studies of policy formation, analyses of voting behavior of legislative, administrative, and judicial officials, and studies of the personal contacts and the quantity and content of communication coming to decision makers' attention. No more than in the previous categories has one single, unifying conceptual framework emerged. It is, therefore, necessary to resort to a paradigmatic presentation of the significant factors which affect decision makers' judgments and acts.

Decision-making theory does not specifically utilize the concept of public opinion but incorporates data on public opinions and policy communications into indexes of perceptual and informational influences upon the perspective of the decision maker. In decision-making vocabulary, "public opinion" means those links or channels of communication with the external setting of the system that are available to the policy maker and the effective rules which define the source and content of communications that are authoritative for him.

"Basic" techniques for analyzing communications linkages and rules (networks) for policy makers are still rudimentary, and to describe the role of public opinion in decision making, we still resort to common-sense terms derived from the "applied" skills of intelligence, propaganda, and public relations. Studies of the impact of public opinion on public policy makers have been focused on the following: (1) policy makers' perceptions of whether a policy proposal or issue would be regarded by the public as necessary, would be tolerated, would be controversial, or would be overwhelmingly unpopular, and the extent to which awareness of such attitudes leads to modifications of policies; (2) policy makers' estimates or anticipations of shifts or changes in public opinion that might result if certain "engineered" events took place, if issues and decisions were presented in certain ways at a certain time, if they were presented by certain prestige or authority figures in

ways and by means recommended by the opinion specialist or public relations consultant (the first category is generally regarded as the specialty of the researcher, the second of the policy adviser on planned programs of "opinion preparation"); and (3) policy makers' acceptance or sharing of the legal and ethical norms held by the majority or at least by substantial numbers of the public.

Policy makers' perceptions of their opinion climate may be rendered more objective by sample surveys. Their estimates of public reactions may be refined by intelligence operations (military, commercial, or political) and employment of the skills of advertising and public relations agencies. Political biographers, journalists, literary artists, and psychological experts have explored the internalized conflicts produced by motivational drives embodying strong compulsions to violate or to conform to societal norms. Whether the context of application be literary, scientific, or practical, empirical analysis supports the assumption of normative theory that public opinion is not a single source of authoritative command; rather, it has at least three different functions: (1) it provides generalized support for the authority and policy-making system; (2) it articulates group interests; and (3) it is a factor in the judgment of individual policy makers.

### Political opinion and public policy

The most pervasive issue dividing theories of the opinion-policy relation bears a striking resemblance to the problem of monism-pluralism in the history of philosophy. The controversy deals with the question of whether the structure of sociopolitical action should be viewed as a more or less centralized process of acts and decisions by a class of key leaders, representing integrated hierarchies of influence in society ("establishment"), or whether it is more accurately envisaged as several sets of relatively autonomous opinion-and-influence groups (based on wealth, votes, prestige, beliefs, or violence), interacting with representative decision makers in an official structure of differentiated (elective, judicial, bureaucratic) governmental authority. The former assumption interprets individual, group, and official action as part of a single system and reduces politics and governmental policies to a derivative of three basic analytical terms: society (social differentiation), culture (norms of expected behavior), and personality (internalized drives and modes of adaptive behavior by leaders). The pluralistic assumption retains the familiar dualism between society (the interpersonal affinities of interest, loyalty, and

opinion exhibited in spontaneous groupings and voluntary associations outside of government) and the state (the official role structures of government, based upon the institutionalized control of violence and bureaucratic performance of public duties). This approach elevates to the central focus of attention the political patterns of communication (superior force, bargaining, competition, coalition, etc.) between leaders of societal groupings and the hierarchies of official policy determination. The monistic assumption is theoretically simpler and more easily verified, particularly in static, rural, local communities. The pluralistic model is more flexible, in that it allows for variations in the configuration of power, depending upon the nature of the issue, and accounts for more of the facts of urban, industrialized, technologically developed national states. Under either assumption, public policy is an end product of different patterns of communication between representative leaders of societal and of governmental opinion.

AVERY LEISERSON

[Directly related are the entries COMMUNICATION, POLITICAL; CROSS PRESSURE; DECISION MAKING; INTERNATIONAL RELATIONS, article on PSYCHOLOGICAL ASPECTS; POLITICAL PROCESS; and the biographies of KEY; LIPPMANN; STOUFFER.]

### BIBLIOGRAPHY

- ALMOND, GABRIEL A. (1950) 1960 *The American People and Foreign Policy*. New York: Praeger.
- BACK, KURT W. 1961 Power, Influence and Pattern of Communication. Pages 137-164 in Luigi Petrullo and Bernard M. Bass (editors), *Leadership and Interpersonal Behavior*. New York: Holt.
- BENDIX, REINHARD 1961 Social Stratification and the Political Community. *Archives européennes de sociologie* 1:181-210.
- BERELSON, BERNARD (1948) 1960 Communications and Public Opinion. Pages 527-543 in Wilbur L. Schramm (editor), *Mass Communications: A Book of Readings*. 2d ed. Urbana: Univ. of Illinois Press.
- CATER, DOUGLASS 1959 *The Fourth Branch of Government*. Boston: Houghton Mifflin.
- DAHL, ROBERT A. (1955) 1956 Hierarchy, Democracy, and Bargaining in Politics and Economics. Pages 83-89 in Heinz Eulau, S. J. Eldersveld, and Morris Janowitz (editors), *Political Behavior: A Reader in Theory and Research*. Glencoe, Ill.: Free Press.
- DOOB, LEONARD W. 1948 *Public Opinion and Propaganda*. New York: Holt.
- KATZ, ELIHU; and LAZARSFELD, PAUL F. (1955) 1964 *Personal Influence: The Part Played by People in the Flow of Mass Communications*. New York: Free Press.
- KEY, V. O. JR. 1961 *Public Opinion and American Democracy*. New York: Knopf.
- KRIS, ERNST; and LEITES, NATHAN C. (1947) 1966 Trends in Twentieth Century Propaganda. Pages 267-277 in Bernard Berelson and Morris Janowitz (edi-



- tors), *Reader in Public Opinion and Communication*. 2d ed. New York: Free Press.
- LANE, ROBERT E. 1959 *Political Life: Why People Get Involved in Politics*. Glencoe, Ill.: Free Press. → A paperback edition was published in 1965.
- LASSWELL, HAROLD D.; and KAPLAN, ABRAHAM 1950 *Power and Society: A Framework for Political Inquiry*. Yale Law School Studies, Vol. 2. New Haven: Yale Univ. Press. → A paperback edition was published in 1963.
- LIKERT, RENSIS 1961 *An Emerging Theory of Organization, Leadership, and Management*. Pages 290–354 in Luigi Petrullo and Bernard M. Bass (editors), *Leadership and Interpersonal Behavior*. New York: Holt.
- LIPSET, SEYMOUR M. 1960 *Political Man: The Social Bases of Politics*. Garden City, N.Y.: Doubleday.
- PETRULLO, LUIGI; and BASS, BERNARD M. (editors) 1961 *Leadership and Interpersonal Behavior*. New York: Holt.
- POLSBY, NELSON W.; DENTLER, R. A.; and SMITH, P. A. (editors) 1963 *Politics and Social Life: An Introduction to Political Behavior*. Boston: Houghton Mifflin.
- ROSENAU, JAMES N. 1961 *Public Opinion and Foreign Policy: An Operational Formula*. New York: Random House.
- SHILS, EDWARD; and JANOWITZ, MORRIS (1948) 1966 *Cohesion and Disintegration in the Wehrmacht in World War II*. Pages 402–417 in Bernard Berelson and Morris Janowitz (editors), *Reader in Public Opinion and Communication*. 2d ed. New York: Free Press. → First published in Volume 12 of the *Public Opinion Quarterly*.
- SNYDER, RICHARD C. (1957) 1964 *A Decision-making Approach to the Study of Political Phenomena*. Pages 3–38 in Roland A. Young (editor), *Approaches to the Study of Politics*. Evanston, Ill.: Northwestern Univ. Press.
- SOCIETY FOR THE PSYCHOLOGICAL STUDY OF SOCIAL ISSUES (1954) 1960 *Public Opinion and Propaganda: A Book of Readings*, by D. Katz et al. New York: Holt. → See especially pages 1–84, "The Nature and Function of Public Opinion," and pages 381–459, "Group Processes: Interaction, Communication and Influence, Social Reinforcement."
- SPIER, HANS 1950 *Historical Development of Public Opinion*. *American Journal of Sociology* 55:376–388.
- TRUMAN, DAVID B. (1951) 1962 *The Governmental Process: Political Interests and Public Opinion*. New York: Knopf.

## PUBLIC POLICY

As a concept in political science, "policy" has at least two quite distinct meanings. On the one hand, "policies" are often considered to be ways of doing things, decision rules—for example: "It is the policy of this office to handle requests in such-and-such a way." In this sense policy would answer the questions "How do you do things around here?" and "What are your rules and procedures?" and may be called administrative policy. Alternatively, policies are often regarded as substantive

programs, referring specifically to the content of what is being done, and not necessarily to how it is being done. Policy in this sense answers the questions "What do you do around here?" and "What kinds of problems do you handle?"

Studies of public policy—the policies of governments—often employ both meanings. Such studies explore not only what agencies of government are doing but how they are doing it, not only the content of the program and its history but its administration as well. The way a program is administered will certainly affect its content and results. Still, the two meanings should be kept distinct, since they refer to analytically separate things. This article will concentrate on the study of the content of governmental activities and will be concerned with administrative policy only insofar as it is studied in conjunction with substantive policy.

The field of public policy has as yet been left relatively untouched by the rigorous scientific investigation that has begun to characterize much of contemporary political science. Although political scientists have spent an enormous amount of time, energy, and resources studying political processes and the contexts in which those processes occur, little attention has been paid to policy studies by those in the discipline whose aim it is to apply scientific criteria to the study of political problems. A major reason for this is the obviously overt value content of public policy questions and the temptation, therefore, of any citizen, professional political scientist or not, to comment on such questions. Indeed, some in the profession would regard the solution of value questions as the major function of the political scientist. But even among those who have such a preference there may be disagreement as to the most fruitful method of proceeding. One could opt for a direct strategy of immediately engaging in value discussion and proposing reforms of one sort or another. Or, alternatively, one could take the position that careful empirical and theoretical analysis is a prerequisite to a proper understanding of the problems.

## Characteristics of policy studies

Although there is some indication of a developing interest in policy analysis by political scientists who are concerned with constructing theory, the most prevalent mode of conducting research on questions of public policy may, by and large, be characterized in the following manner.

**Historical.** Most studies of public policy take as their format a chronological ordering of events, often describing the beginnings and changes in a particular program up to the present time. Break-

ing up the program into convenient historical periods marking major changes in public policy is also a common method of presentation.

The historical approach to public policy is often usefully employed to organize a large body of material, and it provides a logical structure for the discussion of the policies under study. Whether chronological ordering of data is useful for theoretical purposes poses quite a different question. If the historical analysis suggests a trend or pattern of behavior which will then be analyzed and explained, the contribution to theory may be quite large. If, as is more often the case, a historical framework is chosen simply to order the data in some convenient way, then the justification for such an ordering lies elsewhere. The point is that historical discussions of public policy, although undoubtedly informative, may not increase our understanding of the relationship between the policy and other significant factors.

**Descriptive.** By and large, studies of public policy stress richness of descriptive detail. Generalizations about political processes or about comparable public policies are quite rare. Also, in relating our observations of any event we select some features and leave out others, sometimes quite inadvertently. One may legitimately ask, then, of any study, why certain things were described and others not—in other words, what are the criteria employed for the inclusion or exclusion of data? For the most part, in studies of public policies the answer to the criteria question would be "Because it helps to tell the story." But why certain events rather than others? Most researchers, in fact, do have implicit theories which help them to distinguish between the relevant and the irrelevant, the important and the unimportant. To make such criteria explicit is often the beginning of theory and will certainly help the reader decide whether he shares the same generalizations and premises which the author used in choosing events to describe and situations to evaluate. There is no reason why one should have to search for or induce such generalizations from the study. If they are clearly presented by the author, the argument may center on their validity rather than their existence.

**Legal.** As often as not, in policy studies the legal character of the policy is at issue, including its legislative, executive, and judicial history. Other aspects of the political process are sometimes mentioned (such as public opinion and the activities of political parties and interest groups), but the legal descriptive history, including changes in the policy over time, is the more usual framework for policy analyses. Although the legal framework un-

doubtedly acts as a constraint on those involved in the making of public policy, its major effects will be constant for most public policies. Researchers will have to turn to other variables, such as the social, economic, and political setting, the personalities of the actors, and organizational factors, to explain variations within and among public policies.

**Normative.** Most discussions of public policy also have a value position, or an evaluative posture. This view may be either in favor of or in opposition to the current operation of the policy, but the latter is most often the case. In the best of these studies the author will make quite explicit what his value premises are, but occasionally these values will be somewhat hidden and implicit. From this normative position the author will then criticize the program. Often these criticisms will be interlaced in the historical-descriptive-legal narrative, but occasionally they will not appear until the last chapter or two, with possibly some mention of them in the first chapter. The study will then usually conclude with some suggestions for improvement of the program, in some cases with a list of reform measures for "solving" the problem. These suggestions for reform will normally take one of two forms. For some authors they are the *sine qua non* of the study, the major reason why so much time and effort were expended in the collection of historical-legal-descriptive data. For others this reform section is simply *pro forma*; it is expected that traditional public policy studies will have such a section, and leaving it out would slight the expectations of certain readers. Actually, given the relatively straightforward nature of this kind of study, assuming it is competently done, the section on reform may be the only section which is controversial and argumentative. If the facts are correct, the only argument left is in the interpretation of those facts, and since scientific generalization is not one of the stated goals of these studies, value questions fill the gap.

**Other features.** In addition to the above quartet of characteristics, four related observations may be made about policy studies in general.

First, in terms of policy focus, the concern of most writers is with the merits of the policy. Is it a good policy or not? Does it satisfy appropriate values? Is it administered fairly? Is it comprehensive? Are the funds spent too large or too small? By and large, political scientists who are interested in explaining how and why policies develop the way they do have attempted to shy away from these kinds of questions. It is not that these are not questions which all citizens should be asking them-



selves but rather that they are inappropriate questions for a scientific discipline to be asking of its own data.

Second, in terms of approach, the research is problem-oriented, in the sense that the authors speak of problems to be solved with appropriate policies. Again, problem solving of this type is an appropriate form of behavior, but not necessarily for professional political scientists as opposed to political actors. To think in terms of solutions to problems is to direct one's energies away from discovery (the hallmark of the scientific enterprise) and toward a way of doing things which may not be realistically related to the data being studied. How often political scientists propose reforms which are not feasible because they do not square with political realities!

Third, the research is action-oriented in that political scientists doing policy studies often would like to have an impact on the policy to which they address themselves. It is not only their fellow colleagues for whom they write—it is governmental decision makers and opinion leaders as well. It may be more useful, however, for those analyzing public policies to provide as much information as possible on how various aspects of the policy are related to one another, what interests support and oppose the policy, and how the policy is related to the institutions of government and the informal governmental processes, and to leave the activist political role to the decision makers and those whose interests are clearly affected by the policy.

Fourth, in terms of research design, the research is often on a single policy, a case study. The pitfalls of case studies for theory building are well known, but since traditional studies of public policy are problem-oriented and action-oriented, and focus the debate on the merits of specific policies, the case-study format is perfectly appropriate to the aims of the researchers. But for those interested in developing a scientific inquiry, alternative research designs may be more desirable.

Although these observations are meant to be descriptive of most policy studies, there have been departures from this format. For example, there have been several recent efforts to think about the future in certain policy areas. Rather than being concerned with past policy failures, these studies attempt to delineate expected problems sometime in the future, often based on hypothesized trends of certain data. Studies of the population explosion and what to do about it, nuclear strategy, natural and human resources, automation, and other problems have captured the imagination of some researchers. Unlike traditional

studies, they are not as concerned with past or even present problems. Their major concern, rather, is what the world will be like  $x$  number of years hence and possible steps which may be desirable given certain developing problems. However, these studies, too, tend to be problem-oriented single cases differing from most traditional studies primarily in their extensive speculations on the future.

### New directions in policy research

The objectives of science, in any discipline, are relatively straightforward. If one is to have some measure of control over his environment, whether that environment is physical, social, economic, or political, he needs theory that is based on empirical generalizations which, in turn, subsume particular facts. The greater the array of facts subsumed, the more general the theory and the better the understanding of the causal nature of whatever phenomena are being discussed. Science clearly aims at theory construction; and where theory has been developed, manipulation and control—engineering, if you will—have not been far behind. Clearly, the ability of economists, for example, to have as much impact as they do in advising decision makers about the economy is in no small measure due to the development of theories in that field which specify, more or less accurately, the causal mechanisms which influence the economy.

Let us define "theory" as a logically related set of empirical propositions; let us define "empirical proposition" as a statement which relates one variable to another; and let us define "variable" as a concept which may assume different values. If we employ these definitions in the field of public policy, certain considerations are raised in terms of research strategies.

At the minimum we need to transform policy from a constant, an unvarying concept, to a variable. At the present time, for example, it makes little sense to assert that "the more the  $x$  policy, the more the  $y$ ," or "the more the  $y$ , the more the  $x$  policy." "Policy" is simply not a word which has, as yet, achieved the status of a variable.

Traditionally, one possible first step in developing variables is developing typologies, classifications, or taxonomies. A typology (or classification or taxonomy) is a set of concepts or categories which differentiate among various kinds of a higher-order concept. For example, at a very simple level, a typology of a concept like "government" might be "dictatorship" and "democracy." One may then ask a number of questions about

such a typology—for example, whether the categories are exhaustive and/or mutually exclusive.

At a slightly different level, it is also possible to define a typology as a measure of a concept at the nominal level of measurement. When it is viewed as a measurement problem, one can then ask additional questions, such as whether the typology has validity and reliability, and whether operational definitions are easy or difficult to construct.

Once a typology has been constructed, it may then be possible to go further, to inspect whether an underlying factor is present which could move the typology from discrete categories to a continuous variable (increase the measurement power from nominal to ordinal or interval). If one can move in this general direction, it becomes a good deal easier to talk of theory building. However, it would be quite possible to solve all of the problems of measurement, reliability and validity, and inclusiveness and mutual exclusiveness, and still not have contributed much to theory. The next problem to be solved is whether the categories successfully differentiate other phenomena. That is, can the categories be related to some other set of categories in such a way that relationships can be observed which might then achieve the status of theory?

Fortunately, some efforts have been made in this direction, although for the most part in a very primitive way. At a level quite close to the data, researchers have distinguished among policies in terms of their substance (agriculture, labor, education, foreign policy), the institutions making them (legislative, judicial, administrative), the target of the policy (businessmen, Negroes, lower class), time periods in which policy has taken form (ante-bellum, post-depression, prewar, Eisenhower years), ideology (secular, capitalist, first amendment, liberal, conservative), values (good, bad, dangerous), extent of support (consensus, divisive), and governmental level (national, regional, local, metropolitan).

Each of these categories is at a very low level of generalization, and unfortunately very little comparative analysis has taken place within each category. That is, labor policies are seldom contrasted with education policies, national policies are infrequently compared with city policies, etc. Instead, the categories generally are used merely to delineate an area of discourse.

At a somewhat higher level of abstraction, public policies have been differentiated in terms of style versus position issues, symbolic versus material issues; and policies have been categorized

as distributive-regulatory-redistributive, areal-segmental, and collective-private. These distinctions have been useful in developing rather interesting empirical generalizations about how political processes differ, depending upon differences in kinds of policies. For example, it is suggested that political party preference varies more by position issues than by style issues (Berelson et al. 1954, chapter 9); that the interests of organized groups in tangible resources or in substantive power are less easily satiable than are interests in symbolic reassurance, or that the most intensive dissemination of symbols commonly attends the enactment of legislation which is most meaningless in its effects upon resource allocation (Edelman 1964, chapter 2); that political processes themselves will vary, depending upon whether policies attempt to distribute, regulate, or redistribute advantages and disadvantages (Lowi 1964); that policies which affect a whole community are more likely to be successful in homogeneous than in heterogeneous communities, and policies which affect only a segment of a community are more likely to be prevalent in heterogeneous than in homogeneous communities (Froman 1967); and that groups are much more likely to be organized successfully around private than around collective advantages (Olson 1965).

A next step, after developing categories by which different kinds of policies may be compared, is to construct hypotheses in an independent-dependent variable schema. By and large, in the policy literature to date, policies have been treated as dependent variables or outcomes of the political process. For example:

(1) The system of electing the president by the Electoral College tends to favor "liberal" interests.

(2) Apportionment of state legislatures on bases other than population tends to have a "conservative" bias.

(3) Political activities of elites will tend to have a greater pro-civil liberties bias than will political activities of nonelites.

(4) Wealthy, industrial nation-states are more likely to have stable democracies than are poorer, agricultural states.

This use of policy as a dependent variable or outcome need not be its only use, however. It is just as possible to treat policy as an explanation of other phenomena, that is, as an independent variable. One could ask whether different kinds of policies help to structure different types of political processes. At an intuitive level this seems quite plausible, but there has been little or no research into the problem.



Comparative research is also a prerequisite to theory building. Case studies are useful if information about a particular policy is desirable, or if one wishes to explore possible relationships among a wide range of variables. But case studies are manifestly not useful for testing hypotheses or for rigorous theory building. For this, comparative data is essential. Studying two, three, or many policies simultaneously would mean sacrificing factual detail, but it would greatly enlarge the scope of the inquiry and would allow the separation of the unique from the similar.

Once one has constructed variables in a comparative framework, empirical generalization with expanded validity may then take place. These generalizations may be about differences as well as similarities. For example, one may observe that direct federal aid to schools in the United States has, in the last thirty years, proceeded from graduate schools to colleges, and finally to secondary and primary schools. Interestingly enough, racial integration has proceeded in the same order. How does one explain these common patterns? Are there also differences? Are there other policies which exhibit the same pattern? Are there policies which indicate an opposite pattern? If so, how does one explain these differences?

At this point it becomes quite possible to speak of theory construction in the field of public policy. Once variables are identified and placed in an independent-dependent variable framework, comparative data collected and generalization undertaken, placing generalizations in a rigorous logical framework is certainly a possible next step. Given the time and resources which have been trained on political processes in the last twenty years, there is, therefore, no reason to believe that policy analysis cannot assume its rightful place as an important and necessary part of political theory.

LEWIS A. FROMAN, JR.

[See also GOVERNMENT; POLICY SCIENCES; PUBLIC ADMINISTRATION. Specific areas of policy making are discussed in FISCAL POLICY; FOREIGN POLICY; MILITARY POLICY; MONETARY POLICY; NATIONAL SECURITY; PLANNING, ECONOMIC; PLANNING, SOCIAL; WELFARE STATE. Other relevant material may be found in DECISION MAKING; POLITICAL PROCESS.]

#### BIBLIOGRAPHY

- BAUER, RAYMOND A.; POOL, ITHIEL DE SOLA; and DEXTER, LEWIS A. 1963 *American Business and Public Policy*. New York: Atherton.
- BERELSON, BERNARD; LAZARSFELD, PAUL F.; and MCPHEE, WILLIAM N. 1954 *Voting: A Study of Opinion Formation in a Presidential Campaign*. Univ. of Chicago Press.

- BRAYBROOKE, DAVID; and LINDBLOM, CHARLES E. 1963 *A Strategy of Decision: Policy Evaluation as a Social Process*. New York: Free Press.
- BRODIE, BERNARD 1959 *Strategy in the Missile Age*. Princeton Univ. Press.
- DAHL, ROBERT A. (1961) 1963 *Who Governs? Democracy and Power in an American City*. New Haven: Yale Univ. Press.
- EDELMAN, JACOB M. 1964 *The Symbolic Uses of Politics*. Urbana: Univ. of Illinois Press.
- FROMAN, LEWIS A. JR. 1967 An Analysis of Public Policies in Cities. *Journal of Politics* 29:94-108.
- HITCH, CHARLES J.; and MCKEAN, R. N. 1960 *The Economics of Defense in the Nuclear Age*. Cambridge, Mass.: Harvard Univ. Press.
- HUNTINGTON, SAMUEL P. 1961 *The Common Defense: Strategic Programs in National Politics*. New York: Columbia Univ. Press.
- KAHN, HERMAN (1960) 1961 *On Thermonuclear War*. 2d ed. Princeton Univ. Press.
- LIPSET, SEYMOUR M. 1960 *Political Man: The Social Bases of Politics*. Garden City, N.Y.: Doubleday. → A paperback edition was published in 1963.
- LOWI, THEODORE J. 1964 *American Business, Public Policy, Case-studies, and Political Theory*. *World Politics* 16:677-715.
- MUNGER, FRANK J.; and FENNO, RICHARD F. JR. 1962 *National Politics and Federal Aid to Education*. The Economics and Politics of Public Education, No. 3. Syracuse Univ. Press.
- OLSON, MANCUR JR. 1965 *The Logic of Collective Action: Public Goods and the Theory of Groups*. Cambridge, Mass.: Harvard Univ. Press.
- SCHELLING, THOMAS C. 1960 *The Strategy of Conflict*. Cambridge, Mass.: Harvard Univ. Press.
- WILDAVSKY, AARON 1962 *Dixon-Yates: A Study in Power Politics*. Yale Studies in Political Science, Vol. 3. New Haven: Yale Univ. Press.

## PUBLIC RELATIONS

Public relations is the organized effort to communicate information and to modify attitudes and behavior on behalf of a client or a cause. The concept of an organized effort is crucial to this definition, for it points both to the growth in recent decades of a wide assortment of groups seeking to promote some social, political, or economic ideology and to the development of a cadre of specialists whose full-time energies are devoted to winning a greater understanding and acceptance by various publics of an individual, a private business corporation, an agency of government, or other institution.

### History

Such efforts to influence public attitudes are very old indeed, but until the beginning of the twentieth century they were largely carried out in conjunction with other duties and were not defined as constituting a distinct occupation. Poli-

cians gave speeches, travelers wrote accounts of their journeys, religious leaders evangelized, and generals produced memoirs, all without suspecting that they might in subsequent centuries be labeled public relations men. Yet today there are those in the public relations field who claim certain historical figures as early practitioners of their profession: Caesar writing his account of the wars in Gaul; Christ preaching his gospel; Marco Polo writing of his travels to the East, to name but a few. This disarming willingness to pick and choose one's intellectual antecedents is significant chiefly because it reveals a chronic preoccupation of many public relations workers—a sense of insecurity regarding the professional status of their field and a need to legitimatize their work.

Today's public relations practitioner had as one of his earlier ancestors the publicists who made their appearance during the eighteenth century. These writers turned out pamphlets and tracts in support of social, political, and religious causes. But it is important to note that their activities were usually ancillary to other professional interests and almost never provided them with any significant amount of income. More often than not, they donated their services to the group for whom they worked. In England publicists tended to address themselves to the social evils which grew out of the industrial revolution. They inveighed against the shocking conditions of factories employing women and children, or they directed attention to the need for prison reforms or the setting up of mental hospitals and similar institutions to care for the poor and helpless. Similar problems attracted publicists in the United States, but there was, in addition, an overriding concern on the part of many of them during the early decades of the nineteenth century with the abolitionist movement. Following the Civil War some publicists in the United States directed their energies to writing about corruption and abuses of power in the business world and in government. Among these were the muckrakers, who in a sense represented a transitional group from the publicists to the public relations practitioners, for they devoted their full working day to these exposés.

Public relations as a distinct occupation is almost exclusively a development of the twentieth century and the mass society which characterizes it. Big governments, big corporations, and big institutions of all kinds have proliferated since 1900. The managers of these big organizations often feel remote from the groups they serve. Headquarters, whether it be of a government bureau or a business corporation, is often psychologically and physically

distant from its publics. In recent years many managements have employed public relations specialists within their organizations (or as outside consultants) in an effort to overcome hostility, misinformation, or just plain apathy toward themselves.

Business corporations in the United States tend to be larger consumers of public relations services than government bureaus, labor unions, or nonprofit social and voluntary groups. Why should this be so? In the case of government agencies, both state and federal, the party out of power is reluctant to permit too much money to be allocated for information and public relations activities which may enhance the opposition's ability to retain power. Moreover, top elective government officials also rely on their particular political parties to carry out public relations efforts on their behalf. In the case of labor unions and a wide array of nonprofit groups, there is a tradition of voluntary work on the part of members of these organizations at the grass-roots level. The business corporation usually does not command these informal loyalties from its members. Even if it did, there is a question whether management would feel comfortable having its lower echelons initiate public relations actions on its behalf without some kind of strict control of these actions.

### Activities

Since the practice of public relations is most fully developed in the private industrial sector of the United States, it is useful, for illustrative purposes, to detail some of the activities carried out by the public relations department of any large American corporation. With modifications, these operations are representative of the gamut of public relations techniques and programs employed by other institutions and organizations.

Public relations departments of large private industrial companies are expected to provide certain services to their senior executives. They prepare speeches for these corporate officers, assist in answering particularly important correspondence, and provide counsel and guidance to them on a continuing basis regarding the many demands on them to participate in community activities, fund raising, and cultural and social-welfare programs.

Public relations people also have a number of corporate responsibilities. They maintain relations with the press, television, radio, and other media of communications. Requests from these media for information about the company's activities are almost always routed through the public relations department. This department helps to prepare



statements and white papers when its management takes a public stand on an issue involving a government agency or a community problem.

Frequently the public relations department edits a company magazine, which is sent to opinion leaders, shareholders, employees, retailers, and others who are felt to have a particular importance for the firm. It organizes and coordinates film, television, and radio programs on behalf of the company and is charged with developing institutional advertising programs. Institutional advertisements are messages designed primarily to win acceptance of the company per se, rather than specifically aimed at increasing the purchase of its products. (There is obviously a relationship between the kind of product advertising conducted by a company and the impression the public has of that company. In fact, public opinion surveys have found that many people derive their impression of a company from the products it manufactures and the advertising carried out on behalf of those products.)

The public relations department also issues booklets about the company and its activities, annual financial reports, and a variety of specific documents and pictorial material, with the intent of furthering an understanding of the company and its policies. To this end it may also organize exhibits explaining aspects of the company's operations, offer a speakers' program to schools and community groups, and provide tours of manufacturing plants. In the implementation of such programs, public relations departments often single out specific subgroups in the population—educators, clergy, officers of civic and community action groups, and similar elite groups—and develop programs aimed at them.

All the foregoing activities represent the tangible and predictable duties of public relations people. In addition, senior public relations employees and counselors are expected to give continuing advice to their management on present and potential areas of public misunderstanding regarding the company's operations and policies. This may involve mounting a complex and many-faceted campaign directed at coping with a new and serious public relations problem. This is sometimes referred to as "fire fighting." On a more recurrent basis, public relations has the responsibility for the early identification of unfavorable attitudes which may subsequently give trouble. Inasmuch as a public relations department represents but one small segment of a large and frequently complex organization, it must develop sound working relations with other departments.

It is likely to discover that other departments take the view that their company is not primarily organized to win a popularity contest but rather to earn a reasonable profit for its owners. And indeed, there may be occasions when difficult corporate decisions have to be taken in the field of product pricing and wage and salary administration even though they may incur the displeasure of the public. In all institutions, whether profit-making or not, public relations considerations must be weighed in the light of other organizational interests.

### "American" flavor

If, as suggested earlier, public relations is primarily a child of the twentieth century, it is also preponderantly the progeny of the United States of America. A majority of all public relations workers live in the United States, and the field has received much of its impetus and direction from that environment. There is no obvious explanation for the strong American influence in the early growth of public relations. Possibly the extensive saturation by the mass media of communication in the United States facilitated the early development of public relations there. Yet, on a comparative basis, western Europe has always had an intensive spread of those media, especially in the area of the popular press and magazines of opinion; and yet it did not witness a significant growth in public relations until after World War II.

Why does public relations have such a peculiarly American flavor? At least two disparate lines of speculation offer tempting hypotheses. The first sees in the development of public relations an extension of the alleged concern Americans have with being loved and appreciated by others. For example, critics of U.S. foreign policy and information programs have suggested that these efforts are mistakenly directed at winning the affection of other nations rather than their respect. Essayists and journalists who subscribe to the idea that Americans have a pathological need for love might argue that public relations programs on behalf of impersonal institutions merely represent an extension of this need from the individual to the institutional level. Intriguing as this theory might appear, it is based essentially on a facile and tautological assumption regarding the American character which has yet to be documented.

A more plausible theory to account for the early development of public relations in the United States can be based on historical grounds. This would contend that public relations programs are a predictable outgrowth of a very old and respected

American tradition—the democratic belief that all segments of the population should have access to the facts underlying public issues. In such an intellectual context, the emergence of public relations as a distinct field can be seen as a response to the need for more information on a wide variety of political, social, economic, religious, and cultural questions. The relatively simple world of the eighteenth and nineteenth centuries permitted the average U.S. citizen to believe that he could inform himself on the issues of the day. But as America moved into the complex world of the twentieth century, the average citizen simply did not have the time or resources for gathering information on the flood of new problems facing him and his society. Information programs on behalf of special interest groups recognized this need and supplied Americans with material—both pro and con—on issues of interest to them. By contrast, in the more status-conscious and rigid societies of Europe, Latin America, and Asia a similar tradition was lacking, or at least was muted, during the first four decades of the twentieth century. While their citizens may have had the theoretical right to express opinions on a wide variety of public issues, there was far less precedence for them to do so and less demand from them for the information services provided by public relations programs.

### The profession

As a distinct profession, public relations had its beginnings in the early twentieth century, but until the end of World War I there were few persons in the United States or elsewhere whose full-time jobs involved this work. Best known of these was Ivy Lee, whose clients were several of America's largest corporations. But the war, and especially the activities of the United States Committee on Public Information, headed by George Creel, focused public attention on the role that propaganda and information programs play in the struggle for the allegiance of men's minds. The possible commercial application of wartime propaganda techniques was apparent to some executives concerned about public misinformation on business policies. Edward L. Bernays, who had been active in the work of the Committee on Public Information, was a prime mover in the 1920s in developing increased awareness of public relations through a series of successful promotional ventures for important clients. Unhappily, he subsequently identified himself with a phrase, "the engineering of consent," which has reinforced fears in certain intellectual quarters that public relations is essentially a means by which men's minds and

passions may be manipulated on behalf of the vested interests of any client.

During the 1920s and 1930s there was a continued growth in public relations in the United States. Some of the larger advertising agencies took on public relations assignments for their clients. Newspapers and other publications, however, continued to view with suspicion and distaste the "giveaways" of public relations men, feeling that these press releases were efforts to get free advertising. These misgivings lingered in publishing quarters until very recently and contributed to their tendency to equate public relations with publicity. Indeed, the line between public relations and publicity is a blurred one at best, and such distinctions as do exist probably have to do with the time perspectives of these two activities. Publicity seeks to gain maximum and rapid exposure for a client and his program by this exposure per se, as an end in itself. Public relations may also seek wide exposure for a client and his program, but behind this desire for exposure is a detailed program designed to capitalize on the attention won from the public.

Public relations experienced its truly phenomenal growth in the period following World War II. Precise data on its present size have only recently become available because statistics on the number of public relations workers in the United States were not gathered prior to 1960. Before that time the United States Census Bureau classified all public relations workers in the category "editors and reporters." In that year, however, a separate classification was set up—"public relations men and publicity writers"—and 31,141 persons were placed in that category in the 1960 census. Of these, 7,271 were women. The actual number of persons earning a living from jobs involving some aspect of public relations work is probably in excess of 100,000. Public relations workers in the United States tend to cluster in the large metropolitan centers where corporations, trade associations, and government agencies have their headquarters and where the home offices of the mass media are located. For example, in the 1960 census, the New York metropolitan area contained 4,033 of all public relations employees in the United States. The professional association of this field, the Public Relations Society of America, had 5,300 members in 1966.

Outside the United States, the field developed almost exclusively after the end of World War II. The International Public Relations Association listed almost 300 members in 1966. The bulk of these are found in North America and western



Europe. There are, of course, a number of countries where journalists, government information officers, and publicists perform the essential tasks of public relations people but do not choose to take on this title.

**Public relations and the social sciences.** A central assumption of public relations is that its practitioners possess special skills in assessing, interpreting, and modifying public opinion. Committed as they are to understanding the process by which attitudes are formed and changed, they necessarily find themselves working in areas that long have been of theoretical interest to students of social psychology and sociology. One might thus expect that a continuing dialogue would take place between public relations and the social sciences. Such is not the case. In fact, when it takes place at all, this dialogue has tended to be sporadic and relatively barren of results. Several explanations may account for this disappointing state of affairs.

Social scientists, for their part, sometimes profess that public relations problems offer limited intellectual challenge and stimulation. They even suggest that, in actual practice, decisions regarding public relations programs are usually based on intuition rather than on any systematic use of social science knowledge. They may also point out that the field of public relations has made no significant contribution to the body of social science theory or research method, even in the theoretical areas of greatest concern to it: the specifying of variables in the communications process; definition of the role that reference groups play in facilitating or impeding the acceptance of a public relations message; provision of a better description of how the several media of communications are perceived and used by their audiences; and the investigation of the sorting-out process by which audiences choose from the many messages directed at them at any given time.

In terms of professional self-interest, the social scientist may therefore complain that the flow of benefits between his discipline and the public relations world has been pretty much a one-way street. Public relations has adapted findings and has utilized and merchandised theories and jargon from the social sciences for its own profit but has given little of intellectual bounty in return. There is, moreover, a belief on the part of certain social scientists that public relations is willing to manipulate information and attitudes on behalf of any client willing to pay for this expertise, regardless of the social and ethical merits of the client's objectives.

**Criticisms of public relations.** It would be wrong to give the impression that the critics of public relations are drawn solely from academic circles. Novels, films, and other media have frequently pictured the public relations man as a marginal type of operator, interested only in short-term profits and capable of being all things to all people. When these media have chosen not to treat public relations workers in such invidious terms, they have often elected to portray them as comic and impotent figures. Furthermore, within its own ranks, public relations people themselves often evidence an ambivalence and defensiveness about their work.

Little wonder, indeed, that the public relations fraternity has become worried over charges that it lacks professional standards and ethical norms in the conduct of its business. The Public Relations Society of America has taken several steps to mitigate these criticisms. It has adopted a Code of Professional Standards for the Practice of Public Relations and has set up a procedure by which charges of unethical practice can be brought against members of the society; several members have been suspended from the society for violations of this code. In addition, the society has set up a procedure by which its members might be tested and accredited as to their professional competence, and by 1966, 1,100 members had received this accreditation. This procedure for accreditation has evoked mixed reactions from members, some of whom have argued that the practice of public relations is so indeterminate that it defies the drawing up of any rigorous set of standards by which the ability and performance of members might be judged. Arguments on this issue are reminiscent of past wrangles within the ranks of psychologists and sociologists on the feasibility and desirability of certification of members.

Apart from efforts at internal policing of membership, public relations groups have taken steps to reassure the intellectual community in general and the university world in particular regarding their motives and modes of operation. One part of this endeavor has been the setting up of sessions at annual professional meetings of public relations groups for discussion of the relationship of public relations to the humanities and the social sciences. Frequently, distinguished psychologists, sociologists, and anthropologists address these meetings and discuss ways in which better understanding might be promoted between these two worlds. With similar ends in mind, the Public Relations Society of America has sponsored week-long seminars to

provide more intensive exposure for some of its members to the thinking of social scientists and humanists from leading universities. To date, these efforts appear to have served the useful ceremonial function of lessening mutual anxieties and suspicions, but evidence is lacking on whether the day-to-day practices of public relations have been modified because of these interchanges. Certainly a survey of social science journals does not turn up any significant amount of literature reporting on techniques and theories tested in various public relations programs and their possible implications for the study of human behavior.

**Training and research.** Inasmuch as a substantial body of public relations theory rests on an adaptation of research findings and theories from the social sciences, the lack of close liaison between these two worlds is surprising and regrettable. In part, this may be related to the fact that most public relations people are still members of its first generation. They come with a variegated work experience and educational background. A sizable percentage of them are former journalists and free-lance writers, and still others were specialists in the graphic arts, while the rest are former academicians, civil servants, and members of other professions. Many of these public relations workers find social science theory and language remote and apparently impractical in providing answers for the specific, mundane problems of a client. More often than not, the exposure of public relations workers to social science literature has been on a hit-and-miss basis and confined to occasional leisure-time reading, making it easy for them to dismiss social science as the pompous elaboration of the commonplace and obvious. Not all public relations workers are convinced that closer cooperation between themselves and the behavioral scientists is either practical or desirable. For them, public relations demands largely intuitive and creative skills, which do not lend themselves to rigorous scrutiny and categorizing by the fledgling social sciences.

On the other hand, there remains a militant body of public relations experts who are committed to upgrading the professional status of their field, and they seek help from colleges and universities in this quest. Largely as a result of their efforts, over 250 universities in the United States and abroad have set up formal courses of study in public relations, and several have developed graduate programs in which a student may receive an advanced degree in public relations. These latter programs typically require students to spend con-

siderable time studying sociology and social psychology and the specialized areas of communications and public opinion research. As yet, only a tiny fraction of the public relations world has had this training, and it is impossible to comment on the impact it might have on the ethos, outlook, and practices of public relations in the future.

While the social sciences can no longer afford to ignore the field of public relations, by the same token, public relations practitioners must face up to the fact that the intellectual content and theoretical basis of their field have been singularly lacking in exciting new ideas, research practices, or even impressionistic essays on understanding human behavior and the communications process.

### Areas for research

**Effectiveness.** A central problem for public relations is its lack of any convincing yardstick by which to measure its impact and effectiveness. Clients and employers are accustomed to looking for concrete evidence of results for money and effort expended. Product advertising can be related to sales figures and employee relations programs to data on productivity and absenteeism, but public relations efforts are more difficult to evaluate. Countless studies are made each year in which respondents are tested on the degree to which their level of information or attitudes on some issues has been changed as a result of exposure to a public relations program. To date, however, there is a dearth of studies indicating that these professed attitude changes are linked to actual changes in behavior in the marketplace, the polling booth, or in a family or community setting. Lacking such manifest criteria of performance, public relations continues to be defensive about its work.

**Relation to power structure.** A second area calling for greater study is the manner in which the power structure of a community or society affects the reception given to public relations messages. By its own admission, public relations seeks to locate centers of political, social, economic, and intellectual power and apply effective leverage on them. Yet the literature of public relations is excessively modest in recognizing this fact. Public relations texts contain little or no discussion of the classic theories of power as expounded by political scientists, economists, and sociologists.

On many issues facing society there are a variety of possible points of view that can be defended. Clients with conflicting interests and views call upon public relations to help win acceptance



for their positions. A few years ago the public information programs of the trucking industry and the railroads in the United States were in open contention. In the international realm the contest between communist and noncommunist propaganda efforts represents a continuing effort to intellectually disarm and render impotent the competition. On a more mundane level, public relations has been called upon to help product A displace product B from the shelves of stores or to assist in the election of Mr. X over Mr. Y.

Public relations workers are often called upon to take sides in the ordinary as well as in the profoundly important issues of the day. If they do their job well, they may reduce the capacity of other groups in a community to stay alive and sustain their particular economic, social, political, or intellectual positions. Public relations is a force in bringing about social change. Change means that certain ideas, groups, and individuals must move aside and make room for others. Too little recognition has been given in the past to the part public relations plays in modifying the power structure of the society in which it operates.

**Censorship.** Related to the lacuna of data on how the power structure of a society influences the effectiveness of a communications program is the allied topic of censorship—formal and informal. Since public relations programs are designed to create a favorable impression of a client and his cause, facts that might be detrimental to his best interests are often played down or left unmentioned. This is usually justified by saying that one must accentuate the positive and eliminate the negative. The controversy over whether cigarette packages should contain a warning pointing out the health hazards incurred in using them is an example of government agencies and tobacco companies placing very different emphasis on "the facts." In the fields of both domestic and international communications most governments impose some degree of censorship on the information they give out. They frequently come under criticism for being too rigid in the security criteria they use in classifying information. From time to time they must defend themselves against the charge of trying to manage the news. Inasmuch as the word "censorship" has unpleasant connotations, one finds little discussion in the literature on public relations about the ground rules that ought to apply in giving out information on behalf of a client or his cause. Social psychologists have studied the relative effectiveness of presenting one side or both sides of an argument in modifying the attitudes of experimental audiences. Public rela-

tions practitioners have been slow in recognizing that the strategic decisions they make about giving out or withholding information are equally deserving of serious thought and study. All public relations programs operate on some heuristic set of beliefs regarding what may or may not be said to various audiences. Here again, however, little research data are available to tell whether these assumptions are valid. Also lacking are any serious studies of the strategy and rationale for deciding not to communicate in certain public relations situations.

**The role of opinion leaders.** In addition to a study of power in the communications process, there is a need for a more precise definition of the concept of "opinion leader." This term has gained enormous popularity in public relations circles. Whole programs are predicated on reaching certain priority groups or individuals with the client's message and largely ignoring less important ones. Again few studies have been undertaken to judge in retrospect the wisdom and success of designating group A as having a higher priority than group B. In fact, there is limited information and speculation in public relations literature on how the opinion leader exercises his power. Is he merely a gatekeeper who gives the communicator access to important persons? Is his job to embellish, translate, or simplify the communicator's message? Or is the opinion leader's role primarily to serve as an authority figure whose personal endorsement attests to the legitimacy of a message? Campaigns designed to reach opinion leaders would vary a great deal depending on the answers to these questions.

A broader theoretical question, with considerable implications for the field of public relations, is whether these opinion leaders can absorb all the messages directed at them: What happens when hundreds of professional communicators compete for the limited time, attention, and affect of these relatively few people? Do they learn to set up effective defenses against all kinds of information programs? Do they have the time and energy to distinguish between messages on behalf of trivial and significant causes, or are they disposed instead to tune out all efforts to reach them with special pleading? It would be ironic, indeed, if the inordinate efforts by professional communicators to reach these high priority audiences were to result in their becoming immunized to these messages.

**Public opinion research.** The most widely used social science research tool employed in public relations work is the public opinion research study.

In many instances it has provided necessary bench-mark information for measuring the coverage of a public relations effort. However, there are reasons for suspecting that it sometimes is misused in public relations work and that its findings may be misleading. Survey research is most likely to yield fruitful results when the topic under study has rather broad political, social, cultural, or economic implications. If a client's problems fall into these areas, survey research studies will probably make a contribution to improving his fund of information. On the other hand, if the client's prime concern is merely to increase public awareness and acceptance of him, this research method may have limited value. The reason is fairly obvious. Countless business firms, government bureaus, voluntary associations, religious bodies, educational institutions, and community action groups are clamoring for the limited time and interest of the general public. It is totally unrealistic to expect that the average person has more than a most superficial acquaintance or interest in many of these groups. So it happens that a survey that simply asks whether the public knows and approves of certain of these groups is likely to turn up a very high percentage of Don't Know's or No Answer's. Even those who may profess to know these organizations and to hold an opinion about them often turn out, on more intensive investigation, to have only the barest impression of them.

Under these circumstances other social science concepts may offer greater promise for understanding how the public relates to these institutions. For example, it would be useful to know a good deal more about the gratifications and frustrations that people obtain from relating to the impersonal institutions in our society and whether they really expect these institutions to establish an identity and a relationship with them. The literature on personality theory and the socialization of adolescents and adults seems pertinent in such an inquiry.

Implicit in all public relations efforts is the belief that favorable attitudes are desirable and unfavorable attitudes are to be avoided by clients. Why? Do favorable attitudes really get translated into action and how is this done? In some instances, public relations programs seem to supplement marketing and advertising efforts in influencing the sale of a product. In political life, they probably play some role in determining how votes are cast. But the effectiveness of many information and public relations programs is blunted simply because they do not provide the

public with guidance as to what steps it should take to effect the desired change. How often do unfavorable attitudes get translated into actions that are damaging? Every public relations man can cite several horrendous examples of how unfavorable publicity and adverse public reactions led to a drop in sales, the passage of unfavorable legislation, or some other specific action against the offending organization. But these represent specific examples of what happens occasionally when the anger of the public is aroused. We know very little about the significance and possible consequences of a low-grade, unfavorable attitude that persists over time but is not related to any single episode. Does the general public have sufficient involvement in the political process, for example, to translate these simmering disapprovals into action?

From the point of view of a client, may there not be penalties associated with encouraging the public to feel that its voice should be heard in the decisions of management? Does it ever happen that a public relations program becomes too successful and generates pressures from outside groups, which intrude needlessly on the decision-making process of an organization? In fact, does the public truly expect the many large and relatively remote organizations that make up our society to establish rapport with them?

As public relations becomes increasingly a world-wide activity, it is desirable that a careful examination be made of the extent to which its practices and concepts, which have been developed in Western cultures, are applicable to other parts of the world. Public relations programs in Europe and America are predicated on the assumption that their audiences have some level of literacy and free access to the media of communication and that means are available for them to distinguish between the primary and secondary institutions of society. These distinctions may be blurred in cultures where the family and the community take on a far more significant role in the communications network and where an etiquette of listening and responding to strangers may hide the true feeling of an audience.

One of the ironies of public relations today is that it continues to draw from the same small reservoir of academic studies for its intellectual rationale, while all the time having within its own files raw data that might make a significant contribution to communications theory. Simply to list a few topics on which public relations people devote considerable research effort will illustrate the wealth of unexploited data that are collected for



the specific needs of a client but that may be suitable for secondary analysis and publication: pre-test studies of how best to organize messages in order to win the attention of an audience; research on how frequently the same message should be repeated in order to attain optimum impact; and studies of the relationship of the creditability of the communicator and his trade-mark to an audience's predisposition to listen and believe.

Public relations is a new field. Often its own public relations has been poor, especially in dealing with the intellectual community. But in our fragmented and impersonal society, it is being called upon increasingly by a variety of groups to help them communicate more effectively. Its programs influence the lives of many, and for that reason its methods and philosophy deserve careful scrutiny. It represents an important means of developing consensus on problems facing our society, for it permits specialized groups and interests to articulate their points of view. In addition, public relations may help to serve as one of the consciences of a community. Effective programs on behalf of worthy causes make it difficult for anyone to avoid involvement in serious community and national problems. Still another positive function of public relations is that it has legitimized the right of any group to speak out in support of an intellectual position, however unpopular it may be. Responsible citizens are forced to ask themselves whether there may not be two or more sides to a public question. Finally, the practice of public relations has made it necessary for those who communicate on behalf of the special interests of a client to stand up and be counted. By readily identifying the sponsors of a particular message, the public is better able to evaluate the message and the context in which it is proffered.

Public relations continues to be viewed in a narrow and stereotyped way in many quarters. This is unfortunate, for only as the status of public relations is enhanced can it hope to recruit into its ranks persons of superior talent and ethical responsibility.

ROBERT O. CARLSON

[Directly related are the entries ADVERTISING and COMMUNICATION, MASS. Other relevant material may be found in ATTITUDES; PROPAGANDA; PUBLIC OPINION; SURVEY ANALYSIS.]

#### BIBLIOGRAPHY

BAUER, RAYMOND A. 1964 The Obstinate Audience: The Influence Process From the Point of View of Social Communication. *American Psychologist* 19:319-328.

- BARNOW, ERIK 1956 *Mass Communication—Television, Radio, Film, Press: The Media and Their Practice in the United States of America*. New York: Rinehart.
- BERNAYS, EDWARD L. 1923 *Crystallizing Public Opinion*. New York: Boni & Liveright. → Of historical importance as an early effort to state the role of public relations.
- BERNAYS, EDWARD L. 1952 *Public Relations*. Norman: Univ. of Oklahoma Press.
- BERNAYS, EDWARD L. (editor) 1955 *The Engineering of Consent*. Norman: Univ. of Oklahoma Press.
- BURTON, PAUL 1966 *Corporate Public Relations*. New York: Reinhold.
- CARLSON, ROBERT O. 1953 How Can the Social Sciences Meet the Needs of Advertisers? *Printer's Ink* 245:44-47.
- CARLSON, ROBERT O. 1957 The Use of Public Relations Research by Large Corporations. *Public Opinion Quarterly* 21:341-349.
- CARLSON, ROBERT O. 1961 High Noon in the Research Marketplace. *Public Opinion Quarterly* 25:331-341.
- CARLSON, ROBERT O. 1966 The Corporate Manager and the Social Scientist: An Alliance Comes of Age. *Management Review* 55:4-13.
- CENTER, ALLEN H. 1957 *Public Relations Ideas in Action: 500 Tested Public Relations Programs and Techniques*. New York: McGraw-Hill.
- CUTLIP, SCOTT M. (editor) 1966 *A Public Relations Bibliography*. Madison: Univ. of Wisconsin Press.
- CUTLIP, SCOTT M.; and CENTER, ALLEN H. (1952) 1965 *Effective Public Relations*. Englewood Cliffs, N.J.: Prentice-Hall. → A standard textbook.
- DARTNELL CORPORATION (1956) 1961 *The Dartnell Public Relations Handbook*. Edited by John C. Aspley and L. F. van Houten. 3d ed. Chicago: The Corporation.
- FINN, DAVID 1961 The Price of Corporate Vanity. *Harvard Business Review* 39:135-143.
- FOUNDATION FOR RESEARCH ON HUMAN BEHAVIOR, SYMPOSIUM, GOULD HOUSE, 1961 1963 *The Corporation and Its Publics: Essays on the Corporate Image*. Edited by John W. Riley, Jr. New York: Wiley. → The best available discussion by social scientists of the potentials and limits of public relations.
- HARLOW, REX F. 1957 *Social Science in Public Relations: A Survey and an Analysis of Social Science Literature Bearing Upon the Practice of Public Relations*. New York: Harper.
- HARRINGTON, ALAN 1959 *Life in the Crystal Palace*. New York: Knopf. → A whimsical and wise account of one public relations man's view of his work.
- HILL, JOHN W. 1963 *The Making of a Publicity Man*. New York: McKay.
- HYMAN, HERBERT H.; and SHEATSLEY, PAUL B. 1947 Some Reasons Why Information Campaigns Fail. *Public Opinion Quarterly* 11:412-423.
- JACOBSON, HOWARD B. (editor) 1961 *A Mass Communications Dictionary: A Reference Work of Common Terminologies for Press, Print, Broadcast, Film, Advertising, and Communications Research*. New York: Philosophical Library.
- KATZ, ELIHU; and LAZARSFELD, PAUL F. 1955 *Personal Influence: The Part Played by People in the Flow of Mass Communications*. Glencoe, Ill.: Free Press. → An early and important study of informal opinion leaders. A paperback edition was published in 1964.
- KLAPPER, JOSEPH T. 1960 *The Effects of Mass Communication*. Glencoe, Ill.: Free Press. → Summarizes

a wide range of studies regarding the role of the mass media in information programs

- LAZARFELD, PAUL F. 1957 Public Opinion and the Classical Tradition. *Public Opinion Quarterly* 21 39-53
- LEBINGER, OTTO; and SULLIVAN, ALBERT J. 1965 *Information, Influence and Communication: A Reader in Public Relations*. New York: Basic Books.
- LESLEY, PHILIP (editor) (1950) 1962 *Public Relations Handbook*. 2d ed. Englewood Cliffs, N.J.: Prentice-Hall
- MARSTON, JOHN 1963 *The Nature of Public Relations*. New York: McGraw-Hill
- MERTON, ROBERT K. 1946 *Mass Persuasion: The Social Psychology of a War Bond Drive*. New York: Harper. → A detailed sociological analysis of an intensive information program.
- NEWSOM, EARL; CARLSON, ROBERT O.; and KNOTT, LEONARD L. 1959 Feedback: Putting Public Relations in Reverse. Pages 148-162 in Dan J. Fenn, Jr. (editor), *Management's Mission in a New Society*. New York: McGraw-Hill
- NIELANDER, WILLIAM A. 1948 *A Selected and Annotated Bibliography of Literature on Public Relations*. University of Texas, Bureau of Business Research, Bibliography, No. 3. Austin: The Bureau
- PIMLOTT, JOHN A. R. 1951 *Public Relations and American Democracy*. Princeton Univ. Press.
- Public Relations Journal*. → Published since 1945.
- Public Relations News*. → Published since 1944
- Public Relations Quarterly*. → Published since 1937.
- PUBLIC RELATIONS SOCIETY OF AMERICA, INC. 1966 *An Occupational Guide to Public Relations*. New York: The Society.
- ROSS, IRWIN 1959 *The Image Merchants: The Fabulous World of Public Relations*. Garden City, N.Y.: Doubleday
- SAYRE, WILLIAM L. 1963 *The Relations Explosion: A Diagram of the Coming Boom and Shakeout in Corporate Relations*. New York: Macmillan.
- SIMON, RAYMOND (editor) 1966 *Perspectives in Public Relations*. Norman: Univ. of Oklahoma Press
- STEINBERG, CHARLES S. 1958 *The Mass Communicators: Public Relations, Public Opinion, and Mass Media*. New York: Harper
- STEPHENSON, HOWARD (editor) 1960 *Handbook of Public Relations*. New York: McGraw-Hill
- WIEBE, GERNHART D. 1951 Merchandising Commodities and Citizenship on Television. *Public Opinion Quarterly* 15 679-691.

## PUBLIC UTILITIES

See REGULATION OF INDUSTRY; TRANSPORTATION.

## PUNISHMENT

The term "punishment," in its psychological sense, is most commonly and appropriately applied to a situation in which a deprivation or unpleasant experience is deliberately imposed by one party upon another because of an actual or supposed misdeed which is knowingly and intention-

ally committed by the latter. The misdeed may be the violation of a rule, a law, a command, or an expectation and may consist either of an act or of inaction when action is called for. It is generally presumed that both parties to the transaction perceive the "punishment" as unpleasant and the provoking act as a misdeed or at least as punishable. While it is true that innocent persons are sometimes wrongly punished, in principle there can be no such thing as punishment when it is known that there has been no offense. Identical acts may under varying circumstances be perceived as punishment, nonpunishment, or reward by either of the actors involved. The full-fledged punishment situation is thus a complex social relationship in which the perceptions, motives, and intentions of both parties are essential features.

At the sociological level, punishment is characteristically defined in terms of collective perceptions and official definitions. Confusion is likely to follow from the failure to distinguish between the sociological and psychological frames of reference. Thus, from an official point of view, a person who is arrested, tried, and acquitted for an alleged crime has escaped punishment, despite the fact that his reputation may have been destroyed in the process. Punishments which are externally equivalent from the societal standpoint may be very differently perceived by the individuals upon whom they are imposed and may, indeed, not be experienced as punishment at all—as, for example, when an offense is committed to obtain free board and lodging in jail during the cold winter months. Those who regard certain laws as immoral, discriminatory, unjust, or tyrannical often feel that the imposed punishment confers honor and distinction upon them, the question in this case is whether it is the ethical duty of citizens to obey the law or to violate it.

**Extrinsic and intrinsic controls.** Punishments and rewards are generally thought of as contrived or extrinsic consequences of behavior, in contrast to those which are natural or intrinsic effects of a given type of conduct (Lewin 1931). Burning one's finger by touching a flame, being avoided because one is a conversational bore, or suffering indigestion after eating inappropriate food are therefore not ordinarily conceived of as punishment. Punishment is usually applied either to induce a person to do something which he does not wish to do or to refrain from doing what he would like to do. A more satisfactory form of control is achieved by causing people to want to do desirable acts and to avoid undesirable ones, not because of contrived rewards and punishments,



but rather because of satisfactions and dissatisfactions inherent in the behavior. The efficacy of rewards and punishments as measures of social control is far inferior to that exerted by intrinsic factors. It is much more desirable that a child study mathematics because he finds it interesting and believes that it will be useful than that he do so merely because of parental reward or punishment. Organized society presupposes that most members conform to social norms because they find it intrinsically satisfying to do so.

**Unpleasant experiences.** In some theoretical systems, the above distinction between extrinsic and intrinsic consequences of behavior is not made (see Hilgard & Marquis 1940). Thus, experimental psychologists working with lower animals sometimes view any unpleasant experience as punishment. The rat subjected to electric shock in a maze may, for example, be said to be "punished." Sociologists have proposed similar theories. If, for example, interpersonal relationships are conceived, according to the economic model, as exchange relationships in which there are costs and values involved for both parties, punishment may obviously be regarded as a kind of cost. Theories of this sort make it difficult to deal specifically with concepts such as those of responsibility, misdeed, crime, law, guilt, remorse, justice, injustice, conscience, redemption, mercy, and revenge. Indeed, these terms are not ordinarily employed at all by those who use punishment to refer to any unpleasant or painful experience. However, it is generally conceded that conditioning experiments involving "aversive stimulation" of lower animals have only an indirect and peripheral applicability to the human punishment situation, as is indicated by the general tendency to avoid the term "punishment" and to substitute other more neutral and more specific terms.

It is of interest that extensive experimental investigation of the effects of unpleasant experiences upon lower animals, and on human subjects as well, suggests that such experiences are relatively ineffective and unpredictable as a means of behavior control (Hilgard & Marquis 1940). The reasons usually advanced in explanation are that severe pain or punishment tends to elicit strong emotional reactions, such as fear and anxiety, which disrupt behavior patterns and lead to a variety of responses, some of which, especially in human subjects, negate the purposes of punishment. The experimental evidence suggests that the most effective system of rewards and punishments is one in which the former are emphasized and the latter minimized or avoided whenever possible.

**Self-punishment and divine punishment.** Self-punishment and the concept of divine punishment are special cases. Freudian psychologists emphasize that a sense of guilt for a supposed misdeed or sin causes some persons to feel a "need for punishment" and either to punish themselves or to arrange to have others do it for them. In extreme cases, self-punishment may take the form of suicide. It is frequently associated with mental and emotional disorders. The conception of divine punishment may arise from a belief that everything that happens is an expression of God's will and that misfortune, disease, suffering, and death are punishment for sin. If the wicked prosper and the virtuous suffer, it is assumed that the balance will be righted in the life to come.

**The crime of punishment.** The act of punishment in human relations is itself subject to normative control by a wide range of formal and informal rules, the violation of which in some instances transforms the act of punishment into a crime. Parents are generally conceded the right to punish their children for disobedience, but the injury inflicted by a criminal upon his victim for disobedience to his commands is defined as crime, not as punishment. Similarly, excessive parental chastisement of a child may bring formal punishment upon the parent. Critics of current penal practices make frequent use of the phrase "the crime of punishment" (Shaw 1922).

### Formal and informal punishment

Formal punishment is regarded as that which is administered through the courts, and all other kinds are called informal. Informal punishment is of course extensively used in a wide variety of interpersonal and institutional contexts. It is commonly employed in the rearing and education of the young. All types of associations and establishments, such as educational and military institutions, political, occupational, and other types of associations, tend to have established ways of dealing with recalcitrant members. Expulsion from the group is a common sanction in voluntary associations. Punitive sanctions tend to be emphasized when membership is involuntary or in some degree coerced, and they may be specified in relatively formal sets of rules and regulations, as for example in boarding schools, prisons, and other similar establishments. Our concern here, however, will be with punishment for crime.

**Punitive and restitutive sanctions.** While "crime" is defined in different ways by different writers, one may say roughly that it is an act which is viewed as a public rather than a private or per-

sonal wrong and one which is prohibited and made punishable by law. *Criminal* proceedings, which result in the imposition of punitive or repressive sanctions, are contrasted with *civil* proceedings; the latter concern themselves with what may be called restitutive sanctions, such as money or property settlements, which have the effect of compensating or assuaging the injured party and restoring normal relations, much as an apology does in simple person-to-person relationships (Hardy 1963). The wrongs dealt with in civil proceedings are said to be private rather than public, torts rather than crimes; hence, the fine in a criminal case is paid to the state, whereas the damages assessed for a tort go to the aggrieved party. No criminal stigma is supposed to attach to civil proceedings.

There are, however, exceptions and ambiguities in the above distinction. Juvenile courts, for example, are said to be civil in nature, but they produce results which are perceived as punitive by those affected and by the public. The same may be said of compulsory civil commitment as sometimes practiced with respect to sexual psychopaths, alcoholics, drug addicts, and the mentally ill or incompetent. This ambiguity arises largely from the difficulty of determining when deprivation of liberty should be viewed as punishment and when not. The offenders and the general public are little impressed by the doctrine that imprisonment is a means of treatment and rehabilitation rather than of punishment and continue to think that criminal courts punish and that prisons are punitive, even when they are called "therapeutic communities" or "hospitals."

### Processes of formal punishment

The psychological complexities of the punishment relationship, and the variety of its forms, are most strikingly indicated by the elaborate machinery of justice in modern states and by the multiplicity and complexity of the laws, rules, regulations, and principles which control the imposition of punishment. The machinery begins with the legal definition of the kinds of acts deemed to be "criminal" and the selection of those persons who are regarded as "responsible" and therefore fit subjects for court processing; it ends with the release of the offender and the restoration of civil rights. The rules which guide this process are found in constitutions, statutes, and judicial rulings and regulations. In a general way these rules and principles reflect the basic values of the society pertaining to justice, to the nature of the acts for which punishment is appropriate, and to the

kinds of punishment that ought or ought not to be imposed.

**Historical evolution.** The processes of formal punishment in Western civilization have evolved slowly through the centuries and, from a contemporary perspective, appear to have become progressively more complex, refined, humane, discriminating, and flexible. Corporal punishment in all its many manifestations, exile, the death penalty, and all other older kinds of punishment have been largely supplanted by fines and imprisonment. Even where the death penalty is retained, as in most of the states of the United States, it is used less and less. Executions, which were formerly public spectacles and accompanied by torture and mutilation, are now carried out before severely restricted audiences of officials and in what is regarded as the most painless and "humane" manner possible. Torture, which was once popular both as punishment for the guilty and as a means of securing evidence of guilt, is officially outlawed. Officially, punishment for serious crimes today consists overwhelmingly of imprisonment, or loss of liberty. It is true, of course, that the degradation and loss of status associated with conviction for crime may be more painful than the sheer loss of liberty; but this is a factor which the courts cannot control or measure as they do lengths of sentences, and it is not part of the formal punishment.

### Primitive and modern societies

In the modern nation-state, criminal laws can be promulgated and violations punished only by agencies of the state. Some primitive societies have agencies of central political control, and others do not. None of them, of course, have written legal codes and traditions, and in many it is difficult to identify anything like a court or to distinguish law from custom (Malinowski 1926). If it is held that an offense becomes a crime when it is viewed as a menace to the group as a whole and therefore dealt with by representatives of the group, then it might be said that crime and formal punishment emerge in preliterate society when a central agency (such as a chieftain or a council of elders) adjudicates cases and imposes punitive sanctions.

In primitive societies which do not have central governments capable of exercising coercive control over the subgroups which constitute the society, it would not be logical to speak of either crime or formal punishment. The standard example of this type of society is the Ifugao of Luzon (Hoebel 1954). In Ifugao society the kinship groups are sovereign, and most offenses are handled essen-



There are numerous other U. S. Government-owned and -operated facilities that are not included in the survey. The survey is the product of cooperation by the committee on the subject as a whole. The committee is composed of representatives of the various agencies of the Government.

...the ... of ...

... ..

the 1980s, the 1990s, and the 2000s. The 1980s were characterized by a strong emphasis on the development of a national identity and a sense of pride in the country's history and culture. The 1990s saw a shift towards a more global perspective, with a focus on international relations and the role of the United States in the world. The 2000s were marked by a renewed interest in the country's past, with a focus on the role of the military and the importance of national security.

[illegible][illegible]

1977) follows the literature in the sense that the two sets of results are not statistically different. The only significant difference is a low value for the ratio of the average number of observations per site to the number of subjects per site (see Table 1). This is, however, a technical matter and not of substance. In addition, the results are consistent with the findings of the other studies in the literature. For example, the results of the present study are in line with those of the study by Smith and Smith (1976) and the study by Smith and Smith (1977) and the study by Smith and Smith (1978).

[illegible]

The following is the text of the letter from the  
 Secretary of the Board of Directors of the  
 American Red Cross, dated at Washington, D.C.,  
 January 10, 1918, to the Secretary of the  
 War Relocation Authority, at Washington, D.C.:  
 Sir: I have the honor to acknowledge the receipt  
 of your letter of January 8, 1918, and in reply  
 to inform you that the same has been forwarded  
 to the proper authorities for their consideration.  
 Very respectfully,  
 J. H. M. [Signature]

1. The first step is to identify the problem. This involves understanding the current situation and what needs to be changed.



- FAUCONNET, PAUL 1920 *La responsabilité: Étude de sociologie*. Paris: Alcan.
- HALL, JEROME (1947) 1960 *General Principles of Criminal Law*. 2d ed. Indianapolis, Ind.: Bobbs-Merrill.
- HARDY, M. J. L. 1963 *Blood Feuds and the Payment of Blood Money in the Middle East*. Beirut: Privately published.
- HENTIG, HANS VON (1932) 1937 *Punishment, Its Origin, Purpose and Psychology*. London: Hodge. → First published as *Die Strafe: Ursprung, Zweck, Psychologie*.
- HILGARD, ERNEST R.; and MARQUIS, DONALD G. (1940) 1961 *Hilgard and Marquis' Conditioning and Learning*. Revised by Gregory A. Kimble. 2d ed. New York: Appleton. → First published as *Conditioning and Learning*.
- HOEBEL, E. ADAMSON 1954 *The Law of Primitive Man: A Study in Comparative Legal Dynamics*. Cambridge, Mass.: Harvard Univ. Press.
- LEWIN, KURT (1931) 1935 *The Psychological Situations of Reward and Punishment*. Pages 114-170 in Kurt Lewin, *A Dynamic Theory of Personality*. New York: McGraw-Hill.
- MALINOWSKI, BRONISLAW (1926) 1961 *Crime and Custom in Savage Society*. London: Routledge. → A paperback edition was published in 1959 by Littlefield.
- OPPENHEIMER, HEINRICH 1913 *The Rationale of Punishment*. Univ. of London Press.
- RANULF, SVEND (1938) 1964 *Moral Indignation and Middle Class Psychology*. New York: Schocken. → First published in Danish.
- RUSCHE, GEORGE; and KIRCHHEIMER, OTTO 1939 *Punishment and Social Structure*. New York: Columbia Univ. Press.
- SHAW, GEORGE BERNARD (1922) 1946 *The Crime of Imprisonment*. New York: Philosophical Library. → First published as *Imprisonment*.
- SUTHERLAND, EDWIN H.; and CRESSEY, DONALD R. (1924) 1960 *Principles of Criminology*. 6th ed. New York: Lippincott. → First published as a textbook under the title *Criminology*, by Edwin H. Sutherland.

## PURCHASING POWER PARITY

See INTERNATIONAL MONETARY ECONOMICS, article on EXCHANGE RATES.

## PUTNAM, FREDERIC WARD

Frederic Ward Putnam (1839-1915), one of the founders of modern American anthropology, was born in Salem, Massachusetts. He was educated largely by his father, who stimulated the boy's interest in horticulture and in nature generally. This interest led to his election, while still in his teens, to the Essex Institute near his home and to the Boston Society of Natural History. He also met Louis Agassiz, the great Swiss-American naturalist, and under his influence entered the Lawrence Scientific School in 1856, graduating as of 1862. He served as a special assistant under Agassiz until 1864, when he returned to Salem to become curator of vertebrates at the Essex Institute.

Putnam had an early interest in archeological studies. He was among the first to recognize and investigate the archeological remains on the American continent. After holding a number of positions (at the museum of the East India Marine Society in Salem, the museum of the Peabody Academy of Science, and the Penikese Laboratory of the Anderson School of Natural History), Putnam became in 1875 curator of the Peabody Museum of American Archaeology and Ethnology at Harvard University, a position he held until he retired in 1909. He also held a professorship in archeology and ethnology at the university.

From his base at the Peabody Museum, Putnam did much to revolutionize American museum methods by inaugurating scientific expeditions. From 1876 to 1879 he participated in the U.S. geographical and geological surveys west of the 100th meridian under the direction of Lt. G. M. Wheeler; Putnam had charge of the archeological collections made on those surveys and prepared for publication the volume *Archaeology* (1879). He himself directed field research in archeology, especially in New Jersey, Ohio, the southwestern United States, Mexico, Central America, and South America.

In 1891 Putnam was appointed chief of the department of archeology and ethnology for the World's Columbian Exposition in Chicago. The ethnological and natural history collections which he assembled led to the founding of the Columbian Museum in Chicago, now called the Field Museum of Natural History. At the conclusion of his services there he was appointed curator of anthropology at the American Museum of Natural History in New York; one of his assistant curators was Franz Boas. Putnam remained at the American Museum on a part-time basis until 1903, continuing also his administration of the Peabody Museum at Harvard, until his retirement from professional life at the age of 70.

Before he retired, however, he accomplished one more organizational feat for American anthropology. In 1903 he was called to the University of California, where he became professor of anthropology and director of the anthropological museum: this meant that he organized both a museum and a department for the teaching of anthropology at Berkeley.

During his lifetime he published some four hundred papers, reports, and notes in the fields of zoology and anthropology. He also served his profession well as an editor: over a period of 25 years he performed editorial services for the American Association for the Advancement of Science (AAAS) and for all of the museums and departments of

anthropology with which he was affiliated. He was a founder and editor of the *American Naturalist* and the *Naturalists Directory*. After holding the position of secretary of the AAAS for 25 years, he was elected president for 1898-1899.

At Putnam's death, his contributions to anthropology were well described in the *Massachusetts Memorial*: "His skill as an organizer made him the most potent factor in the development of anthropological institutions all over the country . . . his interests centered in the objective, tangible sides of anthropology, and therefore his chief contribution consisted of the development of museum work."

RALPH W. DEXTER

[See also ARCHEOLOGY.]

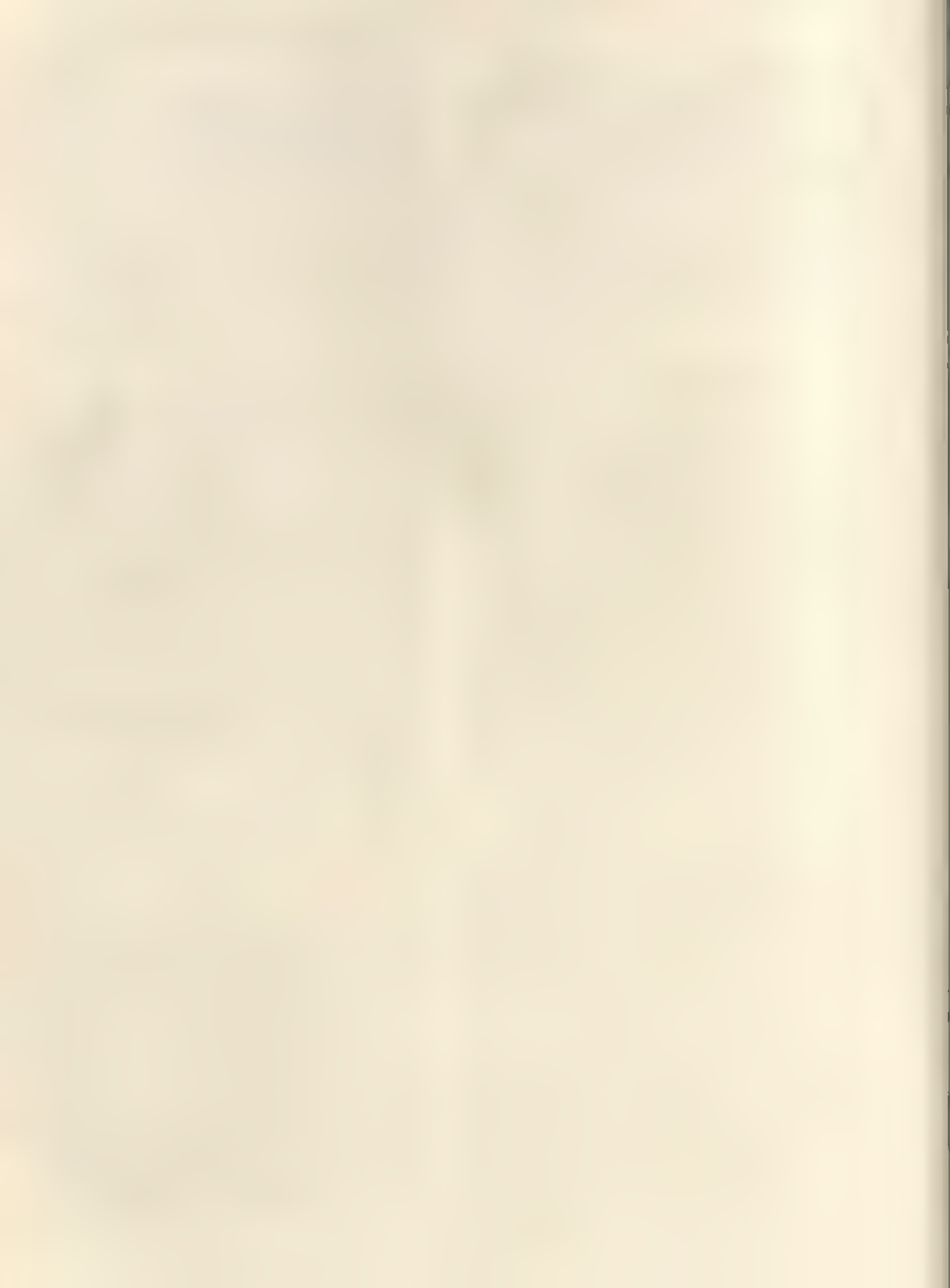
#### WORKS BY PUTNAM

- (1871) 1872 PACKARD, ALPHEUS S.; and PUTNAM, FREDERIC W. *The Mammoth Cave and Its Inhabitants: Or, Descriptions of the Fishes, Insects and Crustaceans Found in the Cave*. Salem, Mass.: Naturalist's Agency.
- 1875a [Archaeological Researches in Kentucky and Indiana.] Boston Society for Natural History, *Proceedings* 17:314-332.
- 1875b [The Swallow Archaeological Collection From New Madrid, Mo.] Harvard University, Peabody Museum of American Archaeology and Ethnology, *Report of the Trustees* 8:16-46. → Reprinted in 1875 in the *American Naturalist* as "The Pottery of the Mound Builders."
- 1878 Archaeological Explorations in Tennessee, 1878. Harvard University, Peabody Museum of American Archaeology and Ethnology, *Report of the Trustees* 11:305-360. → Reprinted in Volume 10 of the *Essex Institute Bulletin*.
- 1879 U.S. GEOGRAPHIC SURVEYS WEST OF THE 100TH MERIDIAN *Report*. Volume 7: Archaeology. Washington: Government Printing Office. → Putnam both edited and contributed to the *Report*.
- 1883a Iron From the Ohio Mounds: A Review of the Statements and Misconceptions of Two Writers of Over Sixty Years Ago. American Antiquarian Society, Worcester, Mass., *Proceedings* New Series 2:349-363.
- 1883b Notes on Copper Implements From Mexico. American Antiquarian Society, Worcester, Mass., *Proceedings* New Series 2:235-246.
- 1887 Notes on the Copper Objects From North and South America, Contained in the Collections of the Peabody Museum. Harvard University, Peabody Museum of American Archaeology and Ethnology, *Report of the Trustees* 15:83-148.
- 1896 PUTNAM, FREDERIC W.; and WILLOUGHBY, C. C. Symbolism in Ancient American Art. American Association for the Advancement of Science, *Proceedings* 44:302-322.

#### SUPPLEMENTARY BIBLIOGRAPHY

- DEXTER, RALPH W. 1964 Frederic Ward Putnam and the Development of Museums of Natural History and Anthropology in the United States. Unpublished manuscript. → This paper was read at the annual meeting of the History of Science Society held in conjunction with the meetings of the American Association for the Advancement of Science at Montreal, Canada, 1964.
- KROEBER, A. L. 1915 Frederic Ward Putnam. *American Anthropologist* New Series 17:712-718.
- MEAD, FRANCES H. 1909 Bibliography of Frederic Ward Putnam. Pages 601-627 in *Putnam Anniversary Volume*. New York: Stechert.
- Putnam Anniversary Volume: Anthropological Essays Presented to Frederic Ward Putnam in Honor of His Seventieth Birthday, April 16, 1909, by His Friends and Associates*. 1909 New York: Stechert.
- TOZZER, A. M. 1935 Frederic Ward Putnam, 1839-1915. Volume 16, pages 125-153 in *National Academy of Sciences, Biographical Memoirs*. Washington: The Academy.







## QUALITATIVE ANALYSIS

*For methods of obtaining, analyzing, and describing qualitative data, see CONTENT ANALYSIS; COUNTED DATA; EVALUATION RESEARCH; FIELD WORK; GRAPHIC PRESENTATION; HISTORIOGRAPHY, especially the article on THE RHETORIC OF HISTORY; INTERVIEWING; OBSERVATION; PERSONALITY MEASUREMENT; PROJECTIVE METHODS; PSYCHOANALYSIS, article on EXPERIMENTAL METHODS; SURVEY ANALYSIS; TABULAR PRESENTATION; TYPOLOGIES.*

## QUALITY CONTROL, STATISTICAL

*"Quality control," in its broadest sense, refers to a spectrum of managerial methods for attempting to maintain the quality of manufactured articles at a desired level. "Statistical quality control" can refer to all those methods that use statistical principles and techniques for the control of quality. In this broad sense, statistical quality control might be regarded as embracing in principle all of statistical methodology. Some areas of statistics, however, have, both historically and as currently used, a special relationship to quality control; it is these areas that are discussed in the articles under this heading. The methods of quality control described in these articles have applications in fields other than industrial manufacturing. For example, process control has been used successfully as an administrative technique in large-scale data-handling organizations, such as census bureaus. Reliability concepts are important in engineering design and seem potentially useful in small-group research.*

- I. ACCEPTANCE SAMPLING
- II. PROCESS CONTROL
- III. RELIABILITY AND LIFE TESTING

H. C. Hamaker  
E. S. Page  
Marvin Zelen

## I ACCEPTANCE SAMPLING

It is generally recognized that mass-production processes inevitably turn out a small amount of product that does not satisfy the specification requirements. Certainly, however, the fraction of defective product should be kept under control, and *acceptance sampling* is one of the methods used for this purpose. Although primarily used in manufacturing situations, the techniques of acceptance sampling can also be applied to nonmanufacturing operations, such as interviewing or editing.

Mass products are usually handled in discrete lots, and these are the units to which acceptance sampling applies. A sample from each lot is inspected, and on the basis of the data provided by the sample it is decided whether the lot as a whole shall be *accepted* or *rejected*.

The actual procedures differ according to the nature of the product and the method of inspection. A distinction must be made between discrete products (nuts and bolts, lamps, radios) and bulk products (coal, fertilizer, liquids). Another important distinction is that between inspection by the method of attributes, where each item inspected is simply classified as defective or nondefective, and inspection by the method of variables, where meaningful numerical measurements on the sample are used to sentence a lot.

The following discussion will mainly concern attribute inspection, which is the technique most widely applied. Variables inspection will be mentioned only briefly.

Systematic investigations of acceptance sampling problems were initiated by two now classic papers of Dodge and Romig (1929-1941), which were later published in book form. Most of the basic



concepts discussed here, such as the AOQL, the LTPD, and single and double sampling, go back to their work.

Further important developments took place during World War II, when acceptance sampling was extensively applied to military supplies. Research was carried out by the Statistical Research Group, Columbia University, and published in two books (1947; 1948).

Since then there has been a steady flow of publications concerned with a variety of both theoretical and practical aspects, such as modifications of sampling procedures, economic principles, the development of sampling standards, or applications in particular situations.

### Method of attributes

**Basic concepts.** Attribute sampling applies to discrete products that are classified as defective or nondefective, regardless of the nature and seriousness of the defects observed. The simplest procedure consists in a *single sampling plan*:

From each lot submitted, a random sample of size  $n$  is inspected. The lot is accepted when the number of defectives,  $x$ , found in the sample is less than or equal to the acceptance number,  $c$ . When  $x$  is greater than or equal to the rejection number,  $c + 1$ , the lot is rejected.

Suppose  $n = 100$  and  $c = 2$ ; that is, 2 per cent defectives are permitted in the sample. This does not mean that a lot containing 2 per cent defectives will always be accepted. Chance fluctuations come into play; sometimes the sample will be better and sometimes worse than the lot as a whole. Statistical theory teaches that if the size of the sample is much smaller than the size of the lot, then a lot with 2 per cent defectives has a *probability of acceptance*  $P_A = 0.68 = 68$  per cent. On the average, from 100 such lots, 68 will be accepted and 32 rejected.

It is clear that the probability of acceptance is a function of the *per cent defective*,  $p$ , in the lot. A plot of  $P_A$  as a function of  $p$  is called the *operating characteristic curve*, or the *O.C. curve*. Every sampling plan has a corresponding O.C. curve, which conveniently portrays its practical performance. The O.C. curves for three single sampling plans are presented in Figure 1.

O.C. curves always have similar shapes, and for practical purposes they can be sufficiently specified by two parameters only. The most important parameters that have been proposed are the following:

(1) *Producer's risk point (PRP)*  $= p_{95}$  = that per cent defective for which the probability of acceptance,  $P_A$ , is 95 per cent.

(2) *Consumer's risk point (CRP)*, more often called the *lot tolerance percentage defective*

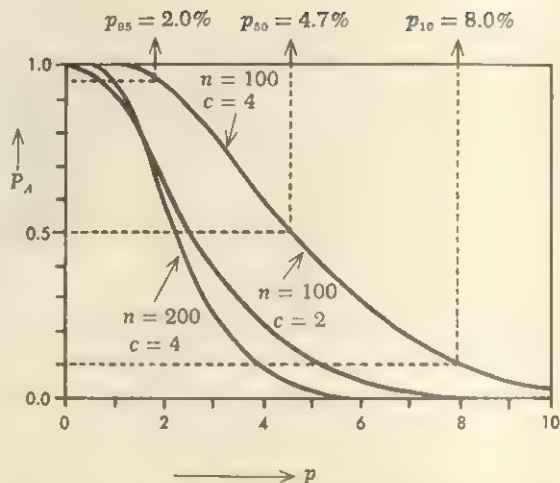


Figure 1 — O.C. curves for three single sampling plans

(LTPD)  $= p_{10}$  = that per cent defective for which  $P_A = 10$  per cent.

(3) *Indifference quality (IQ)* or *point of control*  $= p_{50}$  = that per cent defective for which  $P_A = 50$  per cent.

(4) *Average outgoing quality limit (AOQL)*, which, however, applies only to rectifying inspection, that is, inspection whereby rejected lots are completely screened (100 per cent inspection) and accepted after the removal of all defective items. It can be shown that under rectifying inspection the average outgoing quality (AOQ), that is, the average per cent defective in the accepted lots, can never surpass a certain upper limit, even under the most unfavorable circumstances. This upper limit is the AOQL.

These four parameters are essentially characteristics of the O.C. curves or of the corresponding sampling plans. But the choice of a sampling plan cannot be determined by looking at the O.C. curves alone; in some way the choice must be in keeping with the capabilities of the production process under consideration. Realizing this, Dodge and Romig introduced the *process average*, that is, the average per cent defective found in the samples, as one of the quantities to be used in deciding on a sampling plan. Their sampling plans are such that for lots with a per cent defective equal to the process average the probability of acceptance is always fairly high.

In practice this is not quite satisfactory, because it means that the acceptance sampling system is more lenient toward a poor producer with a high process average than toward a good one with a low average.

Hence the process average has now been super-

sed by the *acceptable quality level* (AQL). This is again a per cent defective and may be considered as an upper limit of acceptability for the process average. Definitions have varied in the course of time. In the earliest version of the *Military Standard 105*, the 105A, which was developed for acceptance sampling of military supplies during World War II and published in 1950, the AQL was defined as a "nominal value to be specified by the U.S. Government." The sampling plans in this standard are so constructed that the probability of acceptance for lots that meet the AQL standard is higher than 90 per cent. This has led other authors to define the AQL instead as the per cent defective with a probability of acceptance of 95 per cent, thus identifying the AQL with the PRP. Thereby, however, the AQL would be essentially a characteristic of a sampling plan and no longer a tolerance requirement of production. The uncertainty seems to have been finally settled by defining the AQL as the maximum per cent defective that, for purposes of acceptance sampling, can be considered satisfactory as a process average. This definition has been adopted in the latest version of the *Military Standard 105*, the 105D, and in a standard on terminology established by the American Society for Quality Control.

Strictly, the O.C. curve of a single sampling plan also depends on the size of the lot,  $N$ , but if, as in most cases, the sample is only a small fraction of the lot, the O.C. curve is almost completely determined by the sample size,  $n$ , and the acceptance number,  $c$ ; the influence of the lot size is unimportant and can be ignored. Since, moreover, the per cent defective is generally very low, a few per cent only, the Poisson formula can be used for computing the O.C. curves [see DISTRIBUTIONS, STATISTICAL, article on SPECIAL DISCRETE DISTRIBUTIONS].

An important logical distinction between contexts must be made: does the O.C. curve relate to the specific lot at hand or to the underlying process that produced the lot? Definitions of consumer's risk, etc., also may be made in either of these contexts. In practice, however, the numerical results are nearly identical, whether one regards as basic a particular lot or the underlying process.

It will be noticed that the O.C. curve of a sampling plan has a close resemblance to the power curve of a one-sided test of a hypothesis [see HYPOTHESIS TESTING]. Indeed, acceptance sampling can be considered as a practical application of hypothesis testing. If on the basis of the findings in the sample the hypothesis  $H_0: p \leq \text{PRP}$  is tested with a significance level of 5 per cent, the lot is rejected if the hypothesis is rejected and the lot is

accepted if the hypothesis is not rejected. Likewise  $H_0: p = \text{PRP}$ ;  $H_1: p = \text{LTPD}$  can be considered as two alternative hypotheses tested against each other with errors of the first and second kind, of 5 per cent and 10 per cent respectively.

*Curtailed, double, multiple, and sequential sampling.* Curtailed sampling means that inspection is stopped as soon as the decision to accept or reject is evident. For example, with a single sampling plan with  $n = 100$ ,  $c = 2$ , inspection can be stopped as soon as a third defective item is observed. This reduces the number of observations without any change in the O.C. curve. The gain is small, however, because bad lots, which are rejected, occur infrequently in normal situations.

Better efficiency is achieved by double sampling plans, which proceed in two stages. First, a sample of size  $n_1$  is inspected, and the lot is accepted if  $x_1 \leq c_1$  and rejected if  $x_1 \geq c_2$ ,  $x_1$  being the number of defectives observed. But when  $c_1 < x_1 < c_2$ , a second sample, of size  $n_2$ , is inspected; and the lot is then finally accepted if the total number of defectives,  $x_1 + x_2$ , is  $\leq c_3$  and is rejected otherwise. The basic idea is that clearly good or bad lots are sentenced by the first sample and only in doubtful cases is a second sample required.

Multiple sampling operates on the same principle but in more than two successive steps, often six or eight. After each step it is decided again whether to accept, reject, or proceed to inspection of another sample. In (fully) sequential sampling, inspection is carried out item by item and a three-way decision taken after each item is inspected.

The basic parameters discussed above apply also to double, multiple, and sequential sampling plans. These procedures can be so constructed that they possess O.C. curves almost identical with that of a given single sampling plan; on the average, however, they require fewer observations. The amount of saving is on the order of 25 per cent for double sampling and up to 50 per cent for multiple and sequential sampling.

Disadvantages of these plans as compared with single sampling plans are more complicated administration and a variable inspection load. Double sampling is fairly often adopted, but multiple and sequential sampling are used only in cases where the cost of inspection is very high, so that a high economy is essential. In many situations single sampling is preferred for the sake of simplicity.

**The choice of a sampling plan.** There are many considerations involved in choosing a sampling plan.

*Sampling standards.* If numerical values are specified for two of the four parameters listed above, the O.C. curve is practically fixed and the



corresponding sampling plan may be derived by computation or from suitable tables. In practice, however, the use of standard sampling tables is much more common.

The best known of these tables is the *Military Standard 105A* (U.S. Department of Defense, Standardization Division 1950), which prescribes a sampling plan in relation to an AQL and the size of the lot; the sample size is made to increase with the lot size so as to reduce the risk of wrong decisions. Three separate tables give single, double, and multiple plans. The user can also choose between different inspection levels, corresponding to systematic changes in all sample sizes; a sample size  $n = 50$  at level I is increased to  $n = 150$  at level III for example.

The *Military Standard 105* also contains rules for a transition to *tightened* or *reduced inspection*. The idea is that, when the sample data indicate that the average per cent defective in the lots received is significantly higher than the required AQL, the consumer should tighten his inspection in order to prevent the acceptance of too many bad lots and to stimulate the producer to improve his production. If, on the other hand, the average per cent defective is significantly lower than the AQL, this indicates that the production process is tightly controlled; inspection is then less essential and the amount of inspection can be reduced.

The *Military Standard 105* has several times been revised; the successive versions are known as the *Military Standard 105A*, *105B*, *105C*, and *105D* (U.S. Department of Defense, Standardization Division 1963). The differences between A, B, and C are relatively unimportant; but in *105D*, issued in 1963, more drastic changes were introduced. These changes were proposed by a special committee of experts jointly appointed by the departments of defense in the United States, Great Britain, and Canada. The basic principles underlying the *Military Standard 105*, however, have not been altered.

The *Military Standard 105* was originally intended for acceptance sampling of military supplies, but it has also been widely applied in industry for other purposes. Other standards have been established by large industrial firms or by governmental offices in other countries. Sometimes these are only modifications of the *Military Standard 105A*; sometimes they have been developed independently and from different points of view. The merits and demerits of a number of these standards are discussed by Hamaker (1960).

*Acceptance sampling as an economic problem.* Many authors have attempted to deduce an opti-

mum sampling plan from economic considerations. Given (a) the cost of inspection of an item, (b) the a priori distribution, that is, the distribution of the per cent defective among the lots submitted for inspection, (c) the loss caused by accepted defectives, and (d) the loss due to rejected nondefectives, one can, in principle, derive an optimum sampling plan, for which the total cost (cost of inspection plus losses) is a minimum.

These economic theories have had little practical success except in isolated cases, mainly, I believe, because their basis is too restricted and because they require detailed information not readily available in industry. The a priori distribution is usually not known and cannot easily be obtained; the consequences of rejecting a lot depend to a very high degree on the stock available and may consequently vary from day to day. Rejection often does not mean that a lot is actually refused; it only means that the inspector cannot accept without consultation with a higher authority or some other authorization.

Since even the most simple products can show a variety of defects, some much more serious than others, it is not easy to see how the loss due to accepted defectives should be estimated. In the *Military Standard 105* this problem has been solved by a classification into critical, major, and minor defects, with a separate AQL for each class. An alternative method, *demerit rating*, consists in scoring points, say 10, 3, and 1, for critical, major, and minor defects and sentencing the lot by the total score resulting from a sample. The advantage is a single judgment instead of three separate judgments, but the theory of O.C. curves no longer applies in a simple and straightforward manner. There are practical problems as to how different defects should be classified and what scores should be assigned to the different classes. These should be solved by discussions between parties interested in the situation envisaged. Some case studies have been described in the literature. No attempts have so far been made to incorporate the classification of defects into economic theories; that would make them yet more complicated and unworkable. This is a decided drawback because in actual practice the degree of seriousness of the defects observed always has a considerable influence on the final decisions concerning the lots inspected. Besides, the economic theories assume that the prevention of accepted defectives is the only purpose of acceptance sampling and this is not correct. It also serves to show the consumer's interest in good quality and to stimulate the producer to take good care of his production processes. Some large firms

have developed vendor rating systems for this very purpose. The information supplied by the samples is systematically collected for each supplier separately and is used in deciding where to place a new contract.

It should not be concluded that no attention is devoted to the economic aspects of the problem of choosing a sampling plan. In acceptance sampling there is always a risk of taking a wrong decision by accepting a bad lot or rejecting a good lot. The larger the lot, the more serious are the economic consequences, but the risk of wrong decisions can be reduced by using larger samples, which give steeper O.C. curves. All existing sampling tables prescribe increased sample sizes with increasing lot sizes, and this practice is derived from economic considerations.

Also, before installing sampling inspection it is always necessary to make some inquiries about the situation to be dealt with. Do bad lots occur, and if so how frequently? How bad are those bad lots? Do accepted defectives cause a lot of trouble? The answers to such questions as these provide crude information about the a priori distribution and the economic aspects, and in choosing an AQL and an inspection level this information is duly taken into account.

Looking at the problem from this point of view, industrial statistics always uses a rough Bayesian approach [see BAYESIAN INFERENCE]. A statistician will never be successful in industry if he does not properly combine his statistical practices with existing technical knowledge and experience and with cost considerations. It is only when an attempt is made to apply the Bayesian principles in a more precise way that problems arise, because the basic parameters required for that purpose cannot easily be estimated with sufficient accuracy. [It is interesting in this connection to compare the empirical Bayesian approach discussed in DECISION THEORY.]

### Method of variables

In sampling by the method of variables a numerical quality characteristic,  $x$ , is measured on each item in the sample. It is usually supposed that  $x$  has a normal distribution [see DISTRIBUTIONS, STATISTICAL] and that a product is defective when  $x$  falls beyond a single specification limit or outside a specification interval: that is, when  $x < L$  and/or  $x > U$ . The basic idea is that from the mean,  $\bar{x}$ , and standard deviation,  $s$ , computed from the sample, an estimate of the per cent defective in the lot can be derived, and this estimate can be compared with the AQL required. For a single

specification limit the criterion operates as follows: If  $x$  is a quality such as length, and the specification requires that not more than a small percentage,  $p$ , of the units in the lot shall have a length greater than  $U$  (upper limit), then the lot is accepted when  $U - \bar{x} \geq ks$  and rejected when  $U - \bar{x} < ks$ , where  $k$  is a constant that depends on  $p$  and the sample size  $n$  and is derived from the theory. For double limits the technique is somewhat more complicated. When the standard deviation is known,  $s$  is replaced by  $\sigma$  and a different value of  $k$  has to be used.

On the basis of earlier suggestions, the theory of sampling by variables was worked out in detail by the Statistical Research Group, Columbia University (1947), and by Bowker and Goode (1952). This theory led to the establishment of the *Military Standard 414* in 1957. This standard has a structure similar to the *Military Standard 105A*; lot size and AQL are the main parameters that determine a sampling plan. Tables are given for one and two specification limits, and for  $\sigma$  both known and unknown.

The advantage of sampling by variables is a more effective use of the information provided by the sample and consequently fewer observations. Where the *Military Standard 105A* prescribes a sample size  $n = 150$ , the *Military Standard 414* uses  $n = 50$  when  $\sigma$  is unknown and  $n$  ranging from 8 to 30 when  $\sigma$  is known.

Disadvantages are that the performance and handling of measurements require more highly trained personnel and that the assumption of a normal distribution is a risky one when it is not known under what circumstances a lot has been produced. For these reasons sampling by variables has found only limited application; sampling by attributes is often preferred.

Present-day reliability requirements have led to an increased interest in life testing procedures and hence to the development of acceptance sampling techniques based on the exponential and the Weibull distributions [see QUALITY CONTROL, STATISTICAL, article on RELIABILITY AND LIFE TESTING].

Acceptance sampling of bulk materials. Acceptance sampling of bulk material constitutes a separate problem. A liquid can be homogenized through stirring, and then the analysis of a single specimen will suffice. Solid material, such as fertilizer or coal, is handled in bales, barrels, wagons, etc. Then, there may be a variability within bales and additional variability among bales. Extensive research is often needed for each product separately before an adequate acceptance sampling



procedure can be developed. A good example is Duncan's investigation of fertilizer (1960).

The theory of sampling by attributes is fairly complete. Research into the economics of acceptance sampling will probably continue for quite a while, but this seems to be an interesting academic exercise that will not lead to drastic changes in industrial practices. A common international sampling standard would be of great practical value, but since in different countries different standards are already established, this is a goal that will not be easily reached.

The theory of sampling by variables may perhaps require some further development. In the *Military Standard 414* the sample sizes prescribed when  $\sigma$  is unknown are three times the size of those for  $\sigma$  known, even for samples as large as 100 or more items. Further research may make possible the reduction of this ratio.

Perhaps the most important new developments will come from new fields of application. It is, for example, recognized that using accountancy to check a financial administration can also be considered as an acceptance sampling procedure and should be dealt with as such. However, the nature of the material and the requirements to be satisfied are entirely different from those in the technological sector. Industrial techniques cannot be taken over without considerable modification, and suitable methods have to be developed afresh. (See Vance & Neter 1956; Trueblood & Cyert 1957.)

H. C. HAMAKER

#### BIBLIOGRAPHY

- BOWKER, ALBERT H.; and GOODE, HENRY P. 1952 *Sampling Inspection by Variables*. New York: McGraw-Hill.
- COLUMBIA UNIVERSITY, STATISTICAL RESEARCH GROUP 1947 *Selected Techniques of Statistical Analysis for Scientific and Industrial Research, and Production and Management Engineering*. New York: McGraw-Hill. → See especially Chapter 1 on the use of variables in acceptance inspection for per cent defective.
- COLUMBIA UNIVERSITY, STATISTICAL RESEARCH GROUP 1948 *Sampling Inspection*. New York: McGraw-Hill. → Describes theory and methods of attribute inspection developed during World War II.
- DEMING, W. EDWARDS 1960 *Sample Design in Business Research*. New York: Wiley. → Contains practical examples of the use of samples in business administration; some of these applications can be considered as problems in acceptance sampling.
- DODGE, HAROLD F.; and ROMIG, HARRY G. (1929-1941) 1959 *Sampling Inspection Tables: Single and Double Sampling*. 2d ed., rev. & enl. New York: Wiley; London: Chapman. → This book is a republication of fundamental papers published by the authors in 1929 and 1941 in the *Bell System Technical Journal*.
- DUNCAN, ACHESON J. (1952) 1965 *Quality Control and Industrial Statistics*. 3d ed. Homewood, Ill.: Irwin. → Contains five chapters on acceptance sampling.
- DUNCAN, ACHESON J. 1960 An Experiment in the Sampling and Analysis of Bagged Fertilizer. *Journal of the Association of Official Agricultural Chemists* 43: 831-904. → A good example of the research needed to establish acceptance-sampling procedures of bulk material.
- GRANT, EUGENE L. (1946) 1964 *Statistical Quality Control*. 3d ed. New York: McGraw-Hill. → The merits and demerits of various acceptance-sampling procedures and standards as well as *Military Standard 105D* are discussed in detail.
- HALD, H. A. 1960 The Compound Hypergeometric Distribution and a System of Single Sampling Inspection Plans Based on Prior Distributions and Costs. *Technometrics* 2:275-340. → Considers sampling from an economic point of view, and contains a fairly complete list of references to earlier literature on this aspect of the sampling problem.
- HAMAKER, HUGO C. 1958 Some Basic Principles of Acceptance Sampling by Attributes. *Applied Statistics* 1:149-159. → Contains a discussion of the difficulties hampering the practical application of economic theories.
- HAMAKER, HUGO C. 1960 Attribute Sampling in Operation. International Statistical Institute. *Bulletin* 37, no. 2:265-281. → A discussion of the merits and demerits of a number of existing sampling standards.
- Quality Control and Applied Statistics: Abstract Service. → Published since 1956. A useful source of information giving fairly complete abstracts of papers published in statistical and technical journals. Of special importance are the classification numbers 200-299: *Sampling Principles and Plans*, and number 823: *Sampling for Reliability*.
- TRUEBLOOD, ROBERT M.; and CYERT, RICHARD M. 1957 *Sampling Techniques in Accounting*. Englewood Cliffs, N.J.: Prentice-Hall.
- U.S. DEPARTMENT OF DEFENSE, STANDARDIZATION DIVISION 1950 *Military Standard 105A*. Washington: Government Printing Office. → A sampling standard widely applied. In 1959 this standard was slightly revised in *Military Standard 105B*. Recently, a further, more drastic revision has been effected jointly by the departments of defense in Canada, Great Britain, and the United States.
- U.S. DEPARTMENT OF DEFENSE, STANDARDIZATION DIVISION 1957 *Military Standard 414*. Washington: Government Printing Office. → A standard for variables inspection corresponding to *Military Standard 105A*.
- U.S. DEPARTMENT OF DEFENSE, STANDARDIZATION DIVISION 1963 *Military Standard 105D*. Washington: Government Printing Office. → Earlier versions of this standard are known as *Military Standard 105A*, *105B*, and *105C*.
- VANCE, LAWRENCE L.; and NETER, JOHN 1956 *Statistical Sampling for Auditors and Accountants*. New York: Wiley.

#### II

#### PROCESS CONTROL

The present article discusses that aspect of statistical quality control relating to the control of a routinely operating process. The traditional and most common field of use is in controlling the

quality of manufactured products, but applications are possible in fields as diverse as learning experiments, stock exchange prices, and error control in the preparation of data for automatic computers. Process control of this kind is usually effected by means of charts that exhibit graphically the temporal behavior of the process; hence, the subject is sometimes called, somewhat superficially, "control charts."

**Inspection for control.** The concept of quality control in its industrial context and the first widely used methods were introduced by W. A. Shewhart (1931). One of the prime concerns was to detect whether the items of output studied had characteristics that behaved like independent observations from a common statistical distribution, that is, whether groups of such items had characteristics behaving like random samples. If the procedure suggested acceptance of the hypothesis of a single distribution, with independence between observations, the production process was said to be "in control" (Shewhart 1931, p. 3), although at this stage the quality of the output, determined by the parameters of the distribution, might not be acceptable. A major contribution was the explicit recognition that such a state of control is necessary before any continuous control of the process can succeed. When the initial examination of the sample data shows the process to be out of control in the above sense, reasons connected with the operation of the process are sought (the search for "assignable causes"; Shewhart 1931, p. 13), improved production methods are introduced, and a further examination is made to see if control has been achieved. Thus, this initial phase of study of a process employs a significance test of a hypothesis in a way encountered in other applications of statistics. It should be noted, however, that often no alternative hypotheses are specified and that, indeed, they are frequently only vaguely realized at the beginning of the investigation.

**Inspection of a process in control.** When the process is in a state of control and its output has the relevant quality characteristics following a distribution with the values of the parameters at their targets (that is, those values of the parameters chosen to cause the output to meet the design specifications with as much tolerance as possible), one task is to determine when the process departs from this state, so that prompt restoring action may be taken. A quality control scheme for this purpose needs, therefore, to give a signal when action is demanded. One of the features of importance for a process inspection scheme is thus the speed at which it detects a change from target.

In conflict with the desire for rapid detection of a change is the necessity for infrequent signals demanding action when no change from target has occurred. The "errors" of signaling a change when none has occurred and of failing to give the signal immediately after a change are similar to the two types of error familiar in hypothesis testing. Whereas in the latter case the probability of such errors is a suitable measure of their occurrence, in quality control the repetition of sampling as the process continues operation and the possibility of combining several samples to decide about a signal make probabilities less convenient measures of error behavior. Instead, the average run length (A.R.L.), defined as the average number of samples taken up to the appearance of a signal, gives a convenient means of comparison for different process inspection schemes (Barnard 1959, p. 240); the A.R.L. is easily related both to the amount of substandard output produced between any change and the signal and to the frequency of unnecessary interference with a controlled process.

*Ancillary tasks of process inspection.* In addition to providing a signal after a change, a process inspection scheme may be required to yield other information. For example, the magnitude of change may need to be estimated so that a dial may be adjusted. Again, coupled with the provision of a signal to take remedial action may be a rule about the destination of any recent production for which there is evidence of a fall in standard, and so the position of any change may need estimation; in this case the scheme is partly one for the deferred sentencing of output, that is, for retrospective acceptance or rejection. In other cases a satisfactory record of a process inspection scheme may guarantee the acceptance of the output by a consumer.

In these and other manufacturing applications where the aim is financial gain, a comparison of alternative schemes in monetary terms should be attempted and not, as so frequently has happened in practice, ignored. Although quantification of many aspects of real situations is difficult or impossible, a monetary comparison can avoid the confusion of using in one situation a measure appropriate for comparing schemes in a totally different situation.

**Types of inspection schemes.** Schemes for the various applications base their rules for the appropriate action to be taken on the results either of single samples of observation or of sequences of such samples; the sequences may be of fixed or variable length. All such schemes are closely related, for those employing a single sample or a fixed number of samples can be exhibited as special



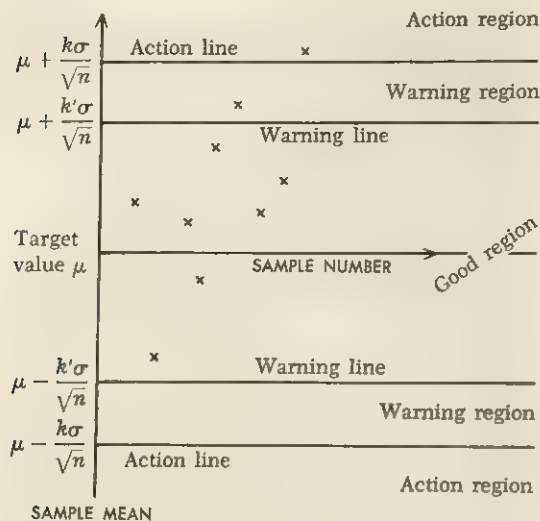


Figure 1 — Chart with warning lines

cases of schemes using a variable number of samples. However, the appearance of the graphical records is quite different. In the first two cases the individual sample results are marked on the chart (Figure 1), while in the last case the sums of all the previous observations, corrected for any expected trend, are plotted (Figure 2). The application of such techniques to different distributions changes the schemes only slightly; the constants involved differ, and sometimes attention is concentrated upon evidence of changes in the distribution parameter in one direction only—for example, increases in the fraction of defective articles produced or in the variance of a measured characteristic are often particularly interesting.

Besides the normal distribution model to be discussed below, another useful model is the Poisson for applications concerning the number of occurrences of a particular type in a given item or length of time, for example, the number of blemishes in sheets of glass of fixed size. The most commonly used schemes are those for controlling the mean

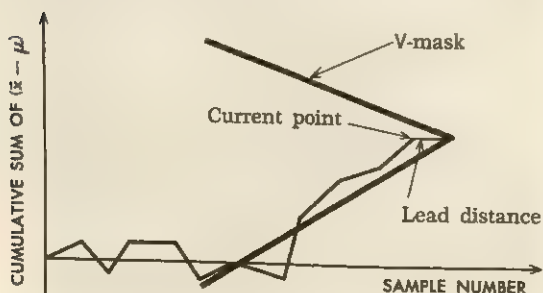


Figure 2 — Cumulative sum chart

of a normal distribution; they are typical of those for other distributions, both continuous and discrete, although schemes for each distribution need separate consideration for calculation of their properties.

The samples used in this sort of activity are usually small, of perhaps four to six observations.

*The Shewhart procedure.* The classical Shewhart procedure for providing a signal when a process in control departed from its target mean  $\mu$  used two "action lines" drawn at  $\mu \pm k\sigma/\sqrt{n}$ , where  $\sigma$  is the process standard deviation estimated from many earlier observations,  $n$  is the size of sample, and  $k$  is a constant chosen pragmatically to be 3 (Shewhart 1931, p. 277) or 3.09 (Dudding & Jennett 1942, p. 64). The signal of a change was received when any sample point, the mean of a sample, fell outside the action lines (Shewhart 1931, p. 290). These rules are such that only about 1 in 500 samples would yield a point outside the action lines if the output were in control with the assumed parameter values; the 1-in-500 choice rests on experience and lacks any foundation from consideration of costs, but the success of these schemes has, in many applications, afforded abundant justification of their use and represented a major advance in process control. For processes where information is required about small changes in the mean, the single small-sample schemes were insufficiently sensitive—small changes need large samples for quick detection—but limitations in sampling effort and a reluctance to sacrifice the possibility of rapid detection of large changes with small samples have caused additions to the original Shewhart scheme.

*Warning line schemes.* Charts of warning line schemes have drawn on them the action lines and two other lines—warning lines—at  $\mu \pm k'\sigma/\sqrt{n}$ , where  $k'$  is a constant less than  $k$  (Dudding & Jennett 1942, p. 14; Grant 1946, art. 159). Accordingly, the chart is divided into action, warning, and good regions (Figure 1). A change in parameter is signaled by rules such as the following: Take action if  $m$  out of the last  $n$  sample points fell in one of the warning regions or if the last point falls in an action region (Page 1955).

Schemes using runs of points are special cases (Grant 1946, art. 88); some that are popular base their action rule on runs of sample points on one side of the target mean (case  $k' = 0$ ,  $k = \infty$ ,  $m = n$ , that is, the action regions disappear and the two warning regions are separated only by a line at the target value: action is taken when a long enough sequence of consecutive points falls in one region).

These schemes retain some of the advantages of small samples and seek to combine the results of

a fixed number of samples in a simple way to increase the sensitivity for sustained small changes in parameter.

**Cumulative sum schemes.** An extension of this idea is provided by the cumulative sum schemes (or cusum schemes), which enjoy the advantages of both large and small samples by combining the relevant information from all recent samples (Barnard 1959, p. 270). Instead of plotting individual sample means,  $\bar{x}$ , on the chart, the differences of these means from the target value,  $\bar{x} - \mu$ , are cumulated and the running total plotted after each sample is taken (Figure 2). A change in process mean causes a change in the direction of the trend of plotted points. One method of defining the conditions for a signal is to place a V-mask on the chart and take action if the arms of the V obscure any of the sample points. The angle of the V and the position of the vertex relative to the last plotted point (the lead distance) can be chosen to achieve the required A.R.L.s. Tables of these constants exist for several important distributions (Ewan & Kemp 1960; Goldsmith & Whitfield 1961; Kemp 1961; Page 1962; 1963). Alternative methods of recording may be adopted for schemes to detect one-sided or two-sided deviations in the parameter. Gauging devices may be used instead of measuring in order to speed the manual operation of such a scheme or to make automatic performance simpler (Page 1962).

**Recent developments.** Attention has been given to schemes for the complete control of the process, that is, procedures for automatically detecting departures from target and making adjustments of appropriate sizes to the control variables (Box & Jenkins 1963; 1964). Naturally, such methods of control are applicable only to those processes for which both measurements of the quality characteristics and the adjustments can be made automatically. Other work has examined different stochastic models of process behavior (Barnard 1959; Bather 1963) and has attempted a comprehensive study of all the costs and savings of a process inspection scheme (Duncan 1956). In industrial applications the financial benefits accruing from the operation of such a scheme are usually of paramount importance, and however difficult it may be to assess them quantitatively, they deserve careful consideration at all stages of the selection and operation of the scheme.

E. S. PAGE

#### BIBLIOGRAPHY

- BARNARD, G. A. 1959 Control Charts and Stochastic Processes. *Journal of the Royal Statistical Society Series B* 21:239-271. → Introduces the V-mask cusum chart and estimation methods.
- BATHER, J. A. 1963 Control Charts and the Minimization of Cost. *Journal of the Royal Statistical Society Series B* 25:49-80.
- BOX, G. E. P., and JENKINS, G. M. 1962 Some Statistical Aspects of Adaptive Optimization and Control. *Journal of the Royal Statistical Society Series B* 24:297-343.
- BOX, G. E. P.; and JENKINS, G. M. 1964 Further Contributions to Adaptive Quality Control; Simultaneous Estimation of Dynamics: Non-zero Costs. *International Statistical Institute, Bulletin* 40:943-974.
- DUDDING, B. P.; and JENNETT, W. J. 1942 *Quality Control Charts*. London: British Standards Institution.
- DUNCAN, ACHESON J. (1952) 1959 *Quality Control and Industrial Statistics*. Rev. ed. Homewood, Ill.: Irwin.
- DUNCAN, ACHESON J. 1956 The Economic Design of  $\bar{x}$  Charts Used to Maintain Current Control of a Process. *Journal of the American Statistical Association* 51:228-242.
- EWAN, W. D.; and KEMP, K. W. 1960 Sampling Inspection of Continuous Processes With No Autocorrelation Between Successive Results. *Biometrika* 47:363-380. → Gives tables and a nomogram for one-sided cusum charts on a normal mean and fraction defective.
- GOLDSMITH, P. L.; and WHITFIELD, H. 1961 Average Run Lengths in Cumulative Chart Quality Control Schemes. *Technometrics* 3:11-20. → Graphs of V-mask schemes for normal means.
- GRANT, EUGENE L. (1946) 1964 *Statistical Quality Control*. 3d ed. New York: McGraw-Hill. → Many examples of the Shewhart chart for different distributions.
- KEMP, K. W. 1961 The Average Run Length of the Cumulative Sum Chart When a V-mask Is Used. *Journal of the Royal Statistical Society Series B* 23:149-153. → Gives tables of cusum schemes for a normal mean.
- PAGE, E. S. 1954 Continuous Inspection Schemes. *Biometrika* 41:100-115. → Introduces cusum schemes.
- PAGE, E. S. 1955 Control Charts With Warning Lines. *Biometrika* 42:242-257.
- PAGE, E. S. 1962 Cumulative Sum Schemes Using Gauging. *Technometrics* 4:97-109.
- PAGE, E. S. 1963 Controlling the Standard Deviation by Cusums and Warning Lines. *Technometrics* 5:307-315. → Gives tables of cusum schemes for a normal range.
- SHEWHART, WALTER A. 1931 *Economic Control of Quality of Manufactured Product*. Princeton, N.J.: Van Nostrand. → The classic volume introducing control chart methods.

#### III

#### RELIABILITY AND LIFE TESTING

Technology has been characterized since the end of World War II by the development of complex systems containing large numbers of subsystems, components, and parts. This trend to even larger and more complex systems is accelerating with the development of space vehicles, electronic computers, and communications and weapons systems. Many of these systems may fail or may operate inefficiently if a single part or component fails. Hence, there is a high premium on having the



...to the ...  
 ...to the ...  
 ...to the ...

to the ...  
 ...to the ...

...to the ...  
 ...to the ...

...to the ...  
 ...to the ...

...to the ...  
 ...to the ...



Figure

...to the ...  
 ...to the ...  
 ...to the ...

...to the ...  
 ...to the ...

...to the ...  
 ...to the ...

...to the ...  
 ...to the ...

...to the ...  
 ...to the ...

...to the ...  
 ...to the ...

...to the ...  
 ...to the ...  
 ...to the ...

...to the ...  
 ...to the ...  
 ...to the ...

...to the ...  
 ...to the ...  
 ...to the ...



Figure 2

and it is important to realize a decreasing in energy, not an increase, is required for the system to self-organize. Therefore, the system must be able to dissipate energy. In the case of a self-organizing system, the energy is dissipated through the system, and the system is able to maintain a steady state. The energy is dissipated through the system, and the system is able to maintain a steady state.

Therefore,

$$\frac{dE}{dt} = -\frac{dE}{dt} \quad (1)$$

which means

$$\frac{dE}{dt} = -\frac{dE}{dt} \quad (2)$$

The system of two resistors shown in Figure 2 is a simple example of a self-organizing system.

Therefore, a self-organizing system is one that is able to maintain a steady state. The system is able to maintain a steady state, and the system is able to maintain a steady state. The system is able to maintain a steady state, and the system is able to maintain a steady state.

$$\frac{dE}{dt} = -\frac{dE}{dt} \quad (3)$$

which means that the system is able to maintain a steady state. The system is able to maintain a steady state, and the system is able to maintain a steady state. The system is able to maintain a steady state, and the system is able to maintain a steady state.

Therefore, a self-organizing system is one that is able to maintain a steady state. The system is able to maintain a steady state, and the system is able to maintain a steady state. The system is able to maintain a steady state, and the system is able to maintain a steady state.

Therefore, a self-organizing system is one that is able to maintain a steady state. The system is able to maintain a steady state, and the system is able to maintain a steady state. The system is able to maintain a steady state, and the system is able to maintain a steady state.

$$\frac{dE}{dt} = -\frac{dE}{dt} \quad (4)$$

which means that the system is able to maintain a steady state. The system is able to maintain a steady state, and the system is able to maintain a steady state. The system is able to maintain a steady state, and the system is able to maintain a steady state.

Therefore, a self-organizing system is one that is able to maintain a steady state. The system is able to maintain a steady state, and the system is able to maintain a steady state. The system is able to maintain a steady state, and the system is able to maintain a steady state.



However, using the two treatment conditions that have the greatest impact on the environment, we can estimate the effect of the two treatment conditions on the environment. The results show that the two treatment conditions have a significant effect on the environment. The results show that the two treatment conditions have a significant effect on the environment. The results show that the two treatment conditions have a significant effect on the environment.

[illegible]





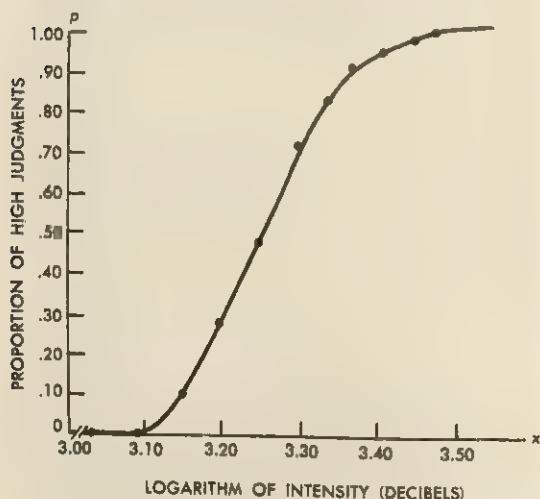
**Table 1 — Proportion of times a sound of varying intensity was judged higher than a standard sound of intensity 1,772 decibels**

Intensity (decibels)	Logarithm of intensity ( $x_i$ )	Proportion of high judgments ( $p_i$ )
1,078	3.03	.00
1,234	3.09	.00
1,402	3.15	.11
1,577	3.20	.28
1,772	3.25	.48
1,972	3.30	.71
2,169	3.34	.83
2,375	3.37	.91
2,579	3.41	.95
2,793	3.45	.98
3,011	3.48	1.00

Source: Data from Spearman 1908.

Figure 1 approximate an underlying smooth sigmoid curve that represents the "true" circumstances.

The value of  $x$  for which the sigmoid curve has value  $\frac{1}{2}$  is taken to be the most important descriptive aspect of the curve by most writers. This half-way value of  $x$ , denoted by *Med*, corresponds to the median if the sigmoid curve is regarded as a cumulative frequency curve. For values of  $x$  below *Med*, the probability of response is less than  $\frac{1}{2}$ ; for values of  $x$  above *Med*, the probability of response is greater than  $\frac{1}{2}$ . Of course, in describing a set of data a distinction must be made between the true *Med* for the underlying smooth curve and an estimate of *Med* from the data. The value *Med* is often called the threshold value in research in sensory perception (see Guilford 1936). This value is not



**Figure 1 — Sigmoid relationship between the logarithm of the intensity of a sound and the proportion of times it is judged higher than a standard sound (intensity 1,772, log intensity 3.25)**

to be confused with the conceptual stimulus value (or increment) that induces *no* response, that is, the 0 per cent response value, sometimes called the absolute threshold. (See Corso 1963 for a review of these concepts.) In pharmacological research the dose at which 50 per cent of the subjects respond is called the E.D. 50 (effective dose fifty), or L.D. 50 (lethal dose fifty) if the response is death.

A natural method of estimating *Med* would be to graph the  $p_i$  against  $x_i$  (as in Figure 1), graduate the data with a smooth line by eye, and estimate *Med* from the graph. In many cases this method is sufficient. The method lacks objectivity, however, and provides no measure of reliability, so that a number of arithmetic procedures have been developed.

**Unweighted least squares.** One early method of estimating *Med* was to fit a straight line to the  $p_i$  by unweighted least squares [see *LINEAR HYPOTHESES, article on REGRESSION*]. This standard procedure leads easily to an estimator of *Med*,

$$\hat{M}_1 = \frac{\frac{1}{2} - \bar{p}}{\hat{B}_1} + \bar{x},$$

where  $\bar{p}$  and  $\bar{x}$  are averages over the levels of  $x$ , and  $\hat{B}_1$ , the estimated slope of the straight line, is

$$\hat{B}_1 = \frac{\sum p_i (x_i - \bar{x})}{\sum (x_i - \bar{x})^2}.$$

**Minimum normit least squares.** The method of unweighted least squares is simple and objective, but it applies simple linear regression to data that usually show a sigmoid trend. Furthermore, the method does not allow for differing amounts of random dispersion in the  $p_i$ .

Empirical and theoretical considerations have suggested the assumption that the expected value of  $p_i$  is related to the  $x_i$  by the sigmoidal function,  $\Phi[(x_i - \text{Med})/\sigma]$ , where  $\sigma$  is the standard deviation of the distribution of the  $x_i$ , and  $\Phi$  denotes the standard normal cumulative distribution. [See *DISTRIBUTIONS, STATISTICAL, article on SPECIAL CONTINUOUS DISTRIBUTIONS*.] Correspondingly, the probability of response at any  $x$  is assumed to be  $\Phi[(x - \text{Med})/\sigma]$ . Notice that  $p = \frac{1}{2}$  at  $x = \text{Med}$ .

The inverse function of  $\Phi$ ,  $Z(p)$ , is called the standard normal deviate of  $p$ , or the normit. To eliminate negative numbers,  $Z(p) + 5$  is used by some and is called the probit. The normit,  $Z(p)$ , is linearly related to  $x$ :

$$(1) \quad Z(p) = -\frac{\text{Med}}{\sigma} + \frac{1}{\sigma} x.$$

If  $x_i$  is plotted against  $Z(p_i)$  from Table 1, a sensibly linear relationship is indeed obtained (see

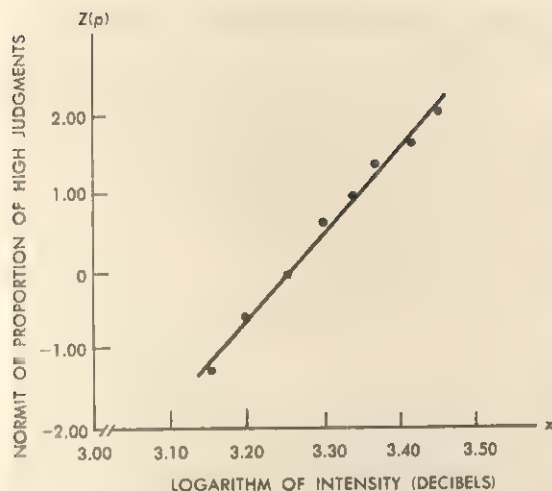


Figure 2 — Linear relationship between the normit,  $Z(p)$ , and the logarithm of sound intensity for Spearman's data

Figure 2), so that the normal assumption seems reasonable for these data.

The minimum normit least squares method (Berkson 1955) is to fit a straight line to the  $Z(p_i)$  by weighted least squares [see LINEAR HYPOTHESES, article on REGRESSION]. The weights are chosen to approximate the reciprocal variances of the  $Z(p_i)$ . Specifically, letting  $y_i = Z(p_i)$  and letting  $z_i$  be the ordinate of the standard normal distribution with cumulative probability  $p_i$ , the weight,  $w_i$ , for  $y_i$  will be

$$w_i = \frac{n_i z_i^2}{p_i(1-p_i)}.$$

The resultant estimator of Med is

$$\hat{M}_2 = -\frac{\bar{y}}{\hat{B}_2} + \bar{x},$$

where  $\bar{x} = \sum w_i x_i / \sum w_i$  and  $\bar{y} = \sum w_i y_i / \sum w_i$  are weighted means and  $\hat{B}_2$  is the estimator of slope,

$$(2) \quad \hat{B}_2 = \frac{\sum w_i y_i (x_i - \bar{x})}{\sum w_i (x_i - \bar{x})^2}.$$

This method is called the minimum normit chi-squared method by Berkson (1955).

**Maximum likelihood estimators.** The method of maximum likelihood, one of general application in statistics, may be applied to quantal response analysis [see ESTIMATION, article on POINT ESTIMATION]. A detailed discussion is given by Finney (1947; 1952). In essence, the technique may be regarded as a weighted least squares procedure, similar to the minimum normit procedure, using

$y_i$  and  $w_i$ , slightly different from the normits and weights of the preceding section. The values are obtained from special tables. The modified weighted least squares procedure is iterated, using new weights stemming from the prior computations, until stability in the estimates is achieved.

**Estimation using the logistic function.** Although the cumulative normal distribution has been most heavily used in quantal response analysis, the logistic function was proposed early and has been advocated for quantal assay in several fields (Baker 1961; Emmens 1940). The function is

$$p = \frac{1}{1 + e^{-(A+Bx)}}.$$

Notice that for  $p = \frac{1}{2}$ ,  $x = \text{Med} = -A/B$ . This function is of sigmoid form and is practically indistinguishable from a cumulative normal, with appropriate choice of  $A$  and  $B$ .

One advantage of the logistic function is that  $x$  can be expressed as a simple function of  $p$ , that is,

$$Y = \ln \frac{p}{1-p} = A + Bx.$$

Thus,  $y_i = \ln p_i / (1-p_i)$  will be approximately linearly related to  $x_i$ . The method of approximate weighted least squares can be used to fit a straight line to the data. The resulting estimator for Med is

$$\hat{M}_3 = -\bar{y} / \hat{B}_3 + \bar{x},$$

where  $\bar{y}$  and  $\bar{x}$  represent weighted means of the  $y_i = \ln [p_i / (1-p_i)]$  and  $x_i$ , with weights  $w_i = n_i p_i (1-p_i)$ , and the estimated slope is

$$\hat{B}_3 = \frac{\sum w_i y_i (x_i - \bar{x})}{\sum w_i (x_i - \bar{x})^2}.$$

The method of maximum likelihood could be applied with the logistic function, but again it proves to be an iterative regression method. However, some simplifications in this case allow the maximum likelihood solution to be graphed; this has been done for certain configurations of  $x$ -levels common in drug assay (Berkson 1960).

**Spearman-Kärber estimation.** Finally, an important method of estimating Med was described by Spearman (1908) and attributed by him to the German psychophysicists. The method is most convenient when the difference between  $x$ -levels is constant, for then the estimator is

$$(3) \quad \hat{M}_4 = x_k + \frac{1}{2}d - d \sum_{i=1}^k p_i,$$

where  $x_k$  is the highest level of  $x$  and  $d$  is the constant difference between levels of  $x$ . Note that an



additional lower level with  $p_0 = 0$  would not change the value of  $\bar{M}_i$ . Neither would an additional higher level of  $x$  with  $p_{k+1} = 1$ , since  $x_k$  would be increased by  $d$ , and an additional  $d$  would be subtracted. A more general formula applicable in the case of unequal spacing of the  $x_i$  is easily available. The expression for  $\bar{M}_i$  above does not require that the group sizes ( $n_i$ ) be equal. The calculation of  $\bar{M}_i$  can be intuitively justified by noting that it is equivalent to reconstructing a histogram from the cumulative distribution formed by graphing the  $p_i$  against the  $x_i$  as a broken line increasing from 0 to 1. The proportion in the histogram between  $x_i$  and  $x_{i+1}$  would be  $p_{i+1} - p_i$ . The mean of the histogram ( $\bar{M}_i$ ) would be the sum of the midpoints of the intervals,  $x_i + \frac{1}{2}d$ , multiplied by their respective relative frequencies,  $p_{i+1} - p_i$ :

$$\bar{M}_i = \sum_{i=0}^k (x_i + \frac{1}{2}d)(p_{i+1} - p_i),$$

where  $p_0 = 0$ ,  $p_{k+1} = 1$ , by definition. It is easily shown that when both are applicable,  $\bar{M}_i = \bar{M}'_i$ , thus demonstrating that the Spearman estimator is simply the mean of a histogram constructed from the quantal response data. Although  $\bar{M}_i$  resembles a mean, it can be regarded as an estimator of the median, Med, for the usual symmetrical sigmoidal curve, since mean and median will be equal in these cases.

The Spearman-Kärber estimator has been criticized because it seems to depend on the possibly fallacious assumption that further  $x$ -levels on either end of the series actually used would have resulted in no responses and all responses, respectively. It should be emphasized here that the heuristic justification above makes use of the unobserved values,  $p_0 = 0$  and  $p_{k+1} = 1$ , but the Spearman-Kärber estimator must be judged on the basis of theory and performance, not on the basis of the heuristic reasoning that suggested the method. Theoretical and practical results suggest that the Spearman-Kärber method has no superior as a general method for estimating Med from quantal data. (See Brown 1961 for a review and for additional results on this point.) Furthermore, the Spearman-Kärber estimator is nonparametric in that no function relating  $p$  and  $x$  appears in the definition of the estimator.

**Other methods.** Other methods of analyzing quantal data are described by Finney (1952). None seems as desirable for general use as the methods described above. The more frequently mentioned procedures are (a) methods based on moving average interpolation of the  $p_i$ , (b) methods based on the angular function as an alternative to the normal or logistic, and (c) the Reed-Muench and

Dragstedt-Behrens method used primarily in biology.

**Sequential estimation of Med.** Occasionally, data on quantal response can be collected most economically in a sequential way and analyzed as they are gathered. As an example, in a clinical trial of a psychotherapeutic drug, individual patients may be allocated to various dose-levels of drug and placebo as they are admitted to, and diagnosed at, a mental hospital. Since the data on treatment effectiveness will become available sequentially and not too rapidly, the evaluation might well be done sequentially. Sequential collection and analysis of quantal response data will yield values of Med (the L.D. 50, or threshold value, etc.) that are more reliable than the fixed-sample-size investigation for comparable numbers of observations. The added precision of the sequential procedure is attained by choosing levels for further observations in the light of the data observed. The result is a concentration of  $x$ -levels in the range where the most information on Med is gained—namely, at  $x$  levels in the vicinity of Med. Several sources (Cochran & Davis 1964; Wetherill 1963) give appropriate methods of carrying out sequential experimentation and analysis specifically for quantal data. [See *SEQUENTIAL ANALYSIS for a general discussion.*]

**Estimation of slope of curve.** All of the preceding discussion of computational procedures has been concerned with computation of an estimator for Med. Another characteristic of the data shown in Figure 1 is the steepness with which the  $p_i$  rise from 0 to 1 as the level of  $x$  is increased. It has been noted that if the function relating  $p$  to  $x$  is regarded as a cumulative frequency function, then Med is the median of the distribution. Similarly, the steepness is related (inversely) to the variability of the frequency function. In particular, if the normal frequency function is used, the slope of the straight line (eq. 1) relating the normal deviate of  $p$  to  $x$  is simply the inverse of the standard deviation of the normal function. Therefore, if the regression methods that stem from least squares or maximum likelihood are used, the resulting value for the slope, for example,  $\hat{B}_2$  in expression (2), could be used to estimate  $\sigma$ . If the logistic functional form is used, the slope can be converted to an estimate of the standard deviation by multiplying the inverse of the slope by a constant,  $\pi/\sqrt{3}$ . The nonparametric Spearman-Kärber estimation of Med does not provide as a side product an estimator of slope or standard deviation. If a value for the standard deviation is desired, a corresponding Spearman procedure for estimating the standard deviation is available (Cornfield & Mantel 1950).

Occasionally, an estimator of a quantile other than the median is of interest. If the tolerance distribution is of normal form, the  $p$ th quantile,  $x_p$ , can be estimated by

$$\hat{x}_p = \hat{M} + z_p \hat{\sigma},$$

where  $\hat{M}$  and  $\hat{\sigma}$  are estimators of the median and standard deviation of the tolerance distribution and  $z_p$  is the  $p$ th quantile of the standard normal distribution.

### Reliability of estimators

The computational procedures described above can be carried out on any set of quantal response data for the purpose of summary or concise description. However, it is apparent that the data are subject to random variation, and this variation, in turn, implies that estimators of Med or  $\sigma$  computed from a given set of quantal response data should be accompanied by standard errors to facilitate proper evaluation. Valid standard errors can be computed only on the basis of a careful examination of the sources and nature of the variation in each specific application, but some widely applicable procedures will be discussed in this section.

**Measuring reliability of the estimators.** The estimators of Med discussed in previous paragraphs will be unbiased, for practical purposes, unless the number of  $x_i$  levels used is small (say, two or three) or the  $x_i$  levels are so widely spaced or poorly chosen that the probability of response is either 0 or 1 at each  $x_i$ . The ideal experiment will have several  $x_i$  levels, with probabilities of response ranging from 5 per cent to 95 per cent. [See ESTIMATION, article on POINT ESTIMATION, for a discussion of unbiasedness.]

A simple method for measuring the reliability of an estimator is to carry out the experiment in several independent replications or complete repetitions. This series of experiments will provide a sequence of estimates for the desired parameter (Med, for example). The mean of the estimates can be taken as the estimator of the parameter, and the standard error of this mean can be computed as the standard deviation of the estimates divided by the square root of the number of estimates. It will be a valid measure of the reliability of the mean.

**Statistical model.** The disadvantage of the procedure described above is that it may not yield error limits as narrow as a method that is tailor-made to the known characteristics of the particular investigation. Furthermore, the statistical properties of the procedure may not be as easily ascertained as for a method based on a more spe-

cific mathematical model. The following model seems to describe quite a number of situations involving quantal response; the estimation procedures discussed above are appropriate to this model.

Take  $P_i$  to be the expected value of  $p_i$  at the level  $x_i$ , with  $n_i$  subjects at this  $i$ th level of  $x_i$  and with  $r_i$  ( $= p_i n_i$ ) of the  $n_i$  subjects responding. Assume that the  $x_i$  are fixed and known without error and that each of the  $n_i$  subjects at  $x_i$  has a probability,  $P_i$ , of responding, independent of all other subjects, whatever their  $x_i$  level. This implies that the observed  $p_i$  are independent, binomially distributed proportions. In particular, each  $p_i$  is an unbiased estimator of its  $P_i$ , with variance  $[P_i(1 - P_i)]/n_i$ .

The assumption of complete independence among all subjects in the investigation must be carefully checked. For example, if the animals at a given dose-level ( $x_i$ ) are littermates, or if the persons at each level of ability ( $x_i$ ) are tested as a group, there may be a serious violation of this assumption, and any assessment of standard error must acknowledge the dependence. (See Finney [1952] 1964, pp. 136-138, for a discussion of procedures appropriate for this type of dependence.) In what follows, complete independence is assumed.

**Standard errors and confidence limits.** In the parametric regression methods, either maximum likelihood or minimum normit chi-squared, estimated standard errors for the estimators of Med and  $\sigma$  are

$$SE_{\hat{\sigma}} = \frac{1}{\hat{\sigma}} \left[ \frac{1}{\sum w_i} + \frac{\hat{M}^2}{\sum w_i (x_i - \bar{x})^2} \right]^{1/2}$$

$$SE_{\hat{M}} = \frac{1}{\hat{\sigma} \sqrt{\sum w_i (x_i - \bar{x})^2}},$$

where  $\hat{\sigma}$ ,  $w_i$ , and  $\bar{x}$  are as defined in previous sections. The same formulas apply for logistic applications. If 95 per cent confidence limits are desired, it is usually satisfactory to take the estimate plus and minus two standard errors. A more exact procedure, developed by Fieller, is discussed in detail by Finney (1952).

The standard error for the nonparametric Spearman procedure is simple and rapidly computed. From expression (3) it can be seen that the Spearman estimator involves only the sum of independent binomially distributed random variables. The usual estimator of standard error is therefore

$$SE_{\hat{d}} = d \sqrt{\sum_{i=1}^h \frac{p_i(1 - p_i)}{n_i}},$$

although there is evidence that it is slightly better to replace  $n_i$  with  $n_i - 1$ . It is useful to note that



the frequency response for a constant stimulus the response is a. the integrated intensity of the action current,  $I$  or  $\Sigma I$ , the point is arrived at before the stimulus (lower rate) is discontinued.

$$m_{\infty} = \int_0^{\infty} \frac{dI}{dt} dt$$

Since it is not an assumption that the point exists, this last equation may be used to find the asymptotic value when the stimulus has been discontinued.

Classical experimental procedure. A short continuous positive stimulus is often used, but this is subject to sufficient variation that more accurate evidence is gained with a constant stimulus, maintained at the asymptotic value some time in advance of the point, when the action potential is about to occur, compared to the very discontinuous stimulus. The method of recording just described is shown in Figure 1.

(1) The stimulus duration was just long enough to make it certain, with an oscilloscope, that the stimulus had reached its peak value.

(2) The stimulus of the integrated current at the time when the maximum stimulus, and the maximum current in the action potential, began.

(3) The onset of specific facilitation, that is, the second response to a second stimulus was about as pronounced as that of the first in the type of stimulus in the case of spontaneous and facilitation responses.

In carrying the frequency-current response  $m_{\infty}$ , it is assumed that the stimulus is constant,  $m_{\infty}$  is constant, and  $\Sigma I$ , or  $I$ , is constant, and is available for some application. The frequency response is used to show a time limit, but not a limit to the time of the peak of stimulus or time of integration time.

#### GROUP RESPONSES IN QUANTAL RESPONSES

Following work in this field, it should be stated that studies of the responses of the quantal response properties have been:

Classical research in synaptic transmission. A short response to a constant stimulus was used to find the summation of a response, and a short stimulus of several integrated responses or several individual responses was used to find the point of facilitation (Davis, et al. 1935). The action was studied in the absence of stimulus. A constant stimulus of facilitation of the action was used to find the point of facilitation of facilitation of facilitation and facilitation of facilitation of facilitation of facilitation.

Classical research in synaptic transmission. A short stimulus of several integrated responses was used to find the summation of a response, and a short stimulus of several integrated responses was used to find the point of facilitation (Davis, et al. 1935). The action was studied in the absence of stimulus. A constant stimulus of facilitation of the action was used to find the point of facilitation of facilitation of facilitation of facilitation.

Classical research in synaptic transmission. A short stimulus of several integrated responses was used to find the summation of a response, and a short stimulus of several integrated responses was used to find the point of facilitation (Davis, et al. 1935). The action was studied in the absence of stimulus. A constant stimulus of facilitation of the action was used to find the point of facilitation of facilitation of facilitation of facilitation.

The present research method is based on the study of the response of the quantal response properties of the action potential, but not of the action potential itself. The action potential is used to find the point of facilitation of the action potential, and the action potential is used to find the point of facilitation of the action potential.

The study of the action potential, and the action potential, is used to find the point of facilitation of the action potential, and the action potential is used to find the point of facilitation of the action potential.

Classical research in synaptic transmission. A short stimulus of several integrated responses was used to find the summation of a response, and a short stimulus of several integrated responses was used to find the point of facilitation (Davis, et al. 1935). The action was studied in the absence of stimulus. A constant stimulus of facilitation of the action was used to find the point of facilitation of facilitation of facilitation of facilitation.

The study of the action potential, and the action potential, is used to find the point of facilitation of the action potential, and the action potential is used to find the point of facilitation of the action potential.





form of proportions of purchasers or of intending purchasers at increasing levels of the continuous variable. Such data could be summarized by estimating the level at which the probability attains some specified value, such as 50 per cent, and the rate at which the probability increases with the continuous variable. This technique could be adapted to the summary of many sample survey collections of data.

The concept of quantal response can be useful in estimating or verifying the validity of an average that might ordinarily be obtained by depending on the memory of persons interviewed. For example, the average age at which a child is able to take several steps without aid might be ascertained by questioning a large group of mothers whose children have been walking for some time. A less obvious, but probably more valid, method of obtaining the estimate would be to interview groups of mothers with children of various ages, ranging from six months to thirty months. The proportion of children walking at each age could be recorded, and this set of quantal response data could be analyzed by the methods discussed above to obtain an estimate of the age at which 50 per cent of the children walk, with a standard error for the estimate. This method has been used to estimate the age of menarche through interviews with adolescent schoolgirls.

BYRON W. BROWN, JR.

[See also COUNTED DATA; PSYCHOPHYSICS.]

#### BIBLIOGRAPHY

- BAKER, FRANK B. 1961 Empirical Comparison of Item Parameters Based on the Logistic and Normal Functions. *Psychometrika* 26:239-246.
- BERKSON, JOSEPH 1955 Estimate of the Integrated Normal Curve by Minimum Normit Chi-square With Particular Reference to Bio-assay. *Journal of the American Statistical Association* 50:529-549.
- BERKSON, JOSEPH 1960 Nomograms for Fitting the Logistic Function by Maximum Likelihood. *Biometrika* 47:121-141.
- BROWN, BYRON W. JR. 1961 Some Properties of the Spearman Estimator in Bioassay. *Biometrika* 48:293-302.
- COCHRAN, WILLIAM G.; and DAVIS, MILES 1964 Stochastic Approximation to the Median Effective Dose in Bioassay. Pages 281-297 in *Stochastic Models in Medicine and Biology: Proceedings of a Symposium* . . . 1963. Edited by John Gurland. Madison: Univ. of Wisconsin Press.
- CORNFIELD, JEROME; and MANTZEL, NATHAN 1950 Some New Aspects of the Application of Maximum Likelihood to the Calculation of the Dosage Response Curve. *Journal of the American Statistical Association* 45: 181-210.
- CORSO, JOHN F. 1963 A Theoretico-Historical Review of the Threshold Concept. *Psychological Bulletin* 60: 356-370.
- EMMENS, C. W. 1940 The Dose/Response Relation for Certain Principles of the Pituitary Gland, and of the Serum and Urine of Pregnancy. *Journal of Endocrinology* 2:194-225.
- FERGUSON, GEORGE A. 1942 Item Selection by the Constant Process. *Psychometrika* 7:19-29.
- FINNEY, DAVID J. (1947) 1962 *Probit Analysis: A Statistical Treatment of the Sigmoid Response Curve*. 2d ed. Cambridge Univ. Press.
- FINNEY, DAVID J. (1952) 1964 *Statistical Method in Biological Assay*. 2d ed. New York: Hafner.
- GUILFORD, JOY P. (1936) 1954 *Psychometric Methods*. 2d ed. New York: McGraw-Hill.
- LAWLEY, D. N. 1943 On Problems Connected With Item Selection and Test Construction. Royal Society of Edinburgh, *Proceedings* 61A:273-287.
- LORD, F. 1952 *A Theory of Test Scores*. Psychometric Monograph No. 7. New York: Psychometric Society.
- MAXWELL, A. E. 1959 Maximum Likelihood Estimates of Item Parameters Using the Logistic Function. *Psychometrika* 24:221-227.
- MILLER, G. A.; and GARNER, W. R. 1944 Effect of Random Presentation on the Psychometric Function: Implications for a Quantal Theory of Discrimination. *American Journal of Psychology* 57:451-467.
- SPEARMAN, C. 1908 The Method of "Right and Wrong Cases" ("Constant Stimuli") Without Gauss's Formulae. *British Journal of Psychology* 2:227-242.
- STEVENS, S. S.; MORGAN, C. T.; and VOLKMANN, J. 1941 Theory of the Neural Quantum in the Discrimination of Loudness and Pitch. *American Journal of Psychology* 54:315-335.
- WETHERILL, G. B. 1963 Sequential Estimation of Quantal Response Curves. *Journal of the Royal Statistical Society Series B* 25:1-48. → Contains 10 pages of discussion by P. Armitage et al.

#### QUASI RENT

See RENT.

#### QUESNAY, FRANÇOIS

François Quesnay (1694-1774), the founder of the economic system that eventually came to be called the physiocratic system, was born in the village of Méré, not far from Versailles. His father, a country lawyer, earned a very small income. Quesnay received no financial aid from his family in securing an education and, indeed, did not learn to read until he was 12. He did, however, somehow acquire a knowledge of medicine, and by 1718, when he was 24 years old, he was practicing medicine in the village of Mantes. Between 1730 and 1758 he wrote several medical treatises that contained original contributions.

Quesnay did not acquire a formal medical degree until 1744, but by 1735 his reputation was such that he obtained a position in Paris as physician to the duke of Villeroy. The next important step in his career was to move to Versailles—under the patronage of Mme. de Pompadour—where he not

only became consulting physician to Louis xv but also came to know the eminent economists of the day. He remained at Versailles thereafter.

Quesnay did not publish his first writing on economics until 1756, when he was in his early sixties; he continued to write on this subject for the rest of his life. Among the several articles he wrote for the *Encyclopédie*, the last two, "Fermiers" (1756) and "Grains" (1757), were his first contributions to economics. In these articles he introduced the view that agriculture, when pursued under favorable conditions, has a unique capacity to yield a *produit net*. The favorable conditions he had in mind were large-scale enterprises backed by abundant capital, such as he had observed during his youth in the wealthy Île-de-France.

Quesnay's ideas long baffled his contemporaries. His *Tableau oeconomique* (1758) contains a comprehensive presentation of his ideas, although it is an extremely brief work. The first edition, consisting of only a few copies, was printed by the king's private press. Rewritten in the form of maxims, it was intended by Quesnay for wider dissemination through publication in an official newspaper. However, Mme. de Pompadour advised him that such publication of a critical view was too risky, and instead Quesnay let his friend the marquis de Mirabeau publish it. Subsequently, he collaborated with the marquis on *Philosophie rurale*, published in Amsterdam in 1763, which turned out to be the clearest statement of the physiocratic doctrine (see Mirabeau 1763).

As a critic of the *ancien régime*, Quesnay wrote for the *Journal de l'agriculture, du commerce et des finances* in 1765 and 1766. Until this journal was banned in 1767, it was a favorite vehicle for expounding physiocratic views. Thereafter, Quesnay and his group established another journal, *Éphémérides du citoyen*, which was entirely devoted to Quesnay's teachings. Taking off from his writings for the *Éphémérides*, Quesnay, at the end of 1767, did some writing for a wider public; with the support of Du Pont de Nemours, this material was published under the title *Physiocratie* (1758-1766). It was this work that provided a name for the physiocratic system of economics and the physiocratic school of economists.

**Bases of the physiocratic system.** The physiocrats assumed that human society is, like the physical universe and analogous to animal societies, subject to natural laws. But whereas, in medieval political philosophy, natural law was simply accepted as the reflection in man of the divine light, or as the participation of a creature endowed with reason in the eternal law, this defini-

tion of natural law clearly no longer sufficed in the age of enlightenment. The physiocrats, in presenting more detailed analyses and explanations of the natural social order, were led to offer a series of propositions that were, in effect, an outline of the basic elements of individualistic capitalism.

According to the physiocratic view, if the constitution of a society is to reflect natural law, then the structure of that society must be determined by the contributions made by its constituent groups toward the production of the national output, and the right of private property and the right to free choice in economic matters must be at the base of the economic system. However, the entire structure must be topped by a strong central power that guarantees to all individuals the enjoyment of the principles of *laissez-faire*—a monarch who holds all political power but whose role is strictly confined to the administration of the natural order of things. Of these elements, only the proposition relating the structure of society to the contributions of component groups was original with the physiocrats. The emphasis on private property and liberty derived largely from John Locke.

Following Locke, the physiocrats justified private property on two grounds: if an object is the product of a man's labor, i.e., if it has been created by its owner; or if the object is the result of the owner's performing one of the three characteristic arts of civilized society. These are the social art (government), the productive art (agriculture), and the sterile art (industry and trade). Similarly, the physiocrats followed Locke in their theory of liberty: liberty is tied to property, and to be free means to be unimpeded in acquiring property and to enjoy what has been legally acquired.

But Locke and his English successors did not integrate their political theory with a systematic account of the over-all productive processes of their society. That a person could acquire a right to property by performing the "social art" would have been incomprehensible to Locke—indeed, a claim so justified would have appeared to him to be a license to usurpation. The originality of the physiocratic doctrine (apart from a series of technical economic points) lay, therefore, in the structuring of a system of political relations upon an economic fundament and the welding of these two parts into an inseparable whole.

**The physiocratic system.** The classes of society that, according to Quesnay, make a contribution to the national product are the "productive" and the "proprietary" classes. Each of these classes is composed of two subgroups: the productive class includes tenant farmers and farm laborers; the



proprietary class includes the landowners and those who exercise political power, pre-eminently the sovereign. Since Quesnay regarded the sovereign as also a large landowner, he somewhat blurred the division within the proprietary class, but his treatment of the political prerogative of the monarch leaves no doubt that he did differentiate between the landowners and the sovereign. His disciples Mercier de la Rivière and the abbé Baudeau consistently made this distinction sharply and precisely.

All four of the above subgroups make contributions toward the maintenance of society, and these contributions justify a certain return. Thus, the monarch and the landowners have a right to their income, the *produit net*, not because of their economic power to exploit the rest of society, nor because of their political privilege, but as a consequence of their contribution to the gross product (*produit brut*). The income of the landowners is justified provided they contribute their landed property to the productive process, that is, if they make the *avances foncières*. Rent, in Quesnay's system, is thus the income landowners receive simply by virtue of owning the only productive agent other than labor. While the net product of labor, whether in agriculture or in industry, is zero, since the laborer consumes as much as he produces, in a "well-regulated kingdom" the net product of land is ordinarily positive. Similarly, the income of the monarch is justified if he makes the *avances souveraines*, that is, if he constructs roads, canals, and other utilities as well as administering the laws which make property secure and maintain the right balance of the economy. The monarch's proper contribution, in short, is to ensure the operation of the natural social order.

The difference between the gross product and the net product represents the income of the tenant farmer and of other persons engaged in agricultural production—the productive class. The tenant farmer also makes a contribution toward the total product: his outlay is composed of two parts, the *avances primitives* and the *avances annuelles*. The former are, roughly speaking, expenditures on the maintenance and replacement of the fixed capital of the farm (fences, farm machinery, etc.); the latter are expenditures on raw materials that are used up annually (seed, manure, etc.) and on the wages of agricultural laborers. The total national product, then, is equal to the sum of the *avances annuelles*, *avances primitives*, and *produit net*.

It is plain from the foregoing why, in Quesnay's view, those who engage in trade and industry are

considered to be the sterile class: this class does not make a contribution to the national product, since it does not use the only factor of production that yields a net surplus—land. Also, the foregoing explains why the physiocrats advocated the concentration of political power in the hands of a monarch. They did not believe in absolutism for its own sake. If they favored a "unique will," they did so because they felt that only concentrated power could overcome the opposition to the beneficent reforms they proposed; only a centralized power could win against separate class interests and establish a social system that would bring the highest public good.

The physiocrats believed that there is only one social system that conforms to the natural order, and its establishment and maintenance must be forever secured against the fickleness of opinion and the arbitrary moods of the masses. Indeed, they feared this arbitrariness of *vox populi* more than the exercise of despotic powers by a just monarch; they realized that the reforms they proposed were rejected because they threatened strongly established privileges and hurt groups of persons whose interests conflicted with the natural order of things. Quesnay and his colleagues attributed the corruption and degeneracy of late eighteenth-century France to the weakness of the monarchy and the prevalent ignorance concerning the true principles of social order, and believed that only legal despotism could save and rebuild France as the divine creator of the natural social order intended.

BERT F. HOSELITZ

[See also ECONOMIC THOUGHT, article on PHYSIOCRATIC THOUGHT; and the biographies of MERCIER DE LA RIVIÈRE; TURCOT.]

#### WORKS BY QUESNAY

- 1756 *Fermiers*. Volume 6, pages 528–540 in *Encyclopédie, ou, dictionnaire raisonné des sciences, des arts et des métiers*. Paris: Briasson.
- 1757 *Grains*. Volume 7, pages 812–831 in *Encyclopédie, ou, dictionnaire raisonné des sciences, des arts et des métiers*. Paris: Briasson.
- (1758) 1766 *The Oeconomical Table: An Attempt Towards Ascertaining and Exhibiting the Source, Progress, and Employment of Riches, With Explanations by the Friend of Mankind, the Celebrated Marquis de Mirabeau*. London: Owen. → First published as *Tableau oeconomique*. Published also as part of Mirabeau's *L'ami des hommes*.
- (1758–1766) 1768 *La physiocratie, ou, constitution naturelle du gouvernement le plus avantageux au genre humain*. 2d ed. 2 vols. Paris and Leiden: Merlin. → A collection of Quesnay's writings. Volume 2 is entitled *Discussions et développemens sur quelques-unes des notions de l'économie politique, pour servir de suite au recueil intitulé: Physiocratie*.

*Oeuvres économiques et philosophiques de F. Quesnay, fondateur du système physiocratique.* Paris: Peelman, 1888.

#### SUPPLEMENTARY BIBLIOGRAPHY

BEER, MAX 1939 *An Inquiry Into Physiocracy.* London: Allen & Unwin.

François Quesnay et la physiocratie. 2 vols. 1958 Paris: Institut National d'Études Démographiques. → Volume 1 consists of essays on Quesnay by various authors, and includes a chronological list of works by Quesnay on pages 301–316 and an annotated bibliography of works about Quesnay on pages 317–392. Volume 2 contains short essays and extracts from his works.

HIGGS, HENRY (1897) 1952 *The Physiocrats: Six Lectures on the French Économistes of the 18th Century.* New York: Langland. → A paperback edition was published in 1963 by Archon.

HISHIYAMA, IZUMI 1960 *The Tableau économique of Quesnay: Its Analysis, Reconstruction, and Application.* *Kyoto University Economic Review* 30, no. 1: 1–46.

MEEK, RONALD L. 1963 *The Economics of Physiocracy: Essays and Translations.* Cambridge, Mass.: Harvard Univ. Press. → The first part includes translations of extracts from various authors. The second part includes four essays by Meek, three of which were previously published between 1951 and 1960.

MIRABEAU, VICTOR DE RIQUETI, DE (1763) 1964 *Philosophie rurale: Ou, économie générale et politique de l'agriculture, réduite à l'ordre immuable des lois physiques et morales, qui assurent la prospérité des empires.* 2d ed. 3 vols. Amsterdam: Libraires Associés. → Chapter 7, "Les rapports des dépenses entre elles," was written by Quesnay.

SCHELLE, GUSTAVE 1907 *Le docteur Quesnay: Chirurgien, médecin de Madame de Pompadour et de Louis XV, physiocrate.* Paris: Alcan.

### QUETELET, ADOLPHE

Lambert Adolphe Jacques Quetelet (1796–1874), best known for his contributions to statistics, was born in the Belgian city of Ghent. When he was seven his father died, and Quetelet, on finishing secondary school at the age of 17, was forced to earn his own living. He accepted a post as teacher of mathematics in a secondary school at Ghent, but his true inclination at the time was toward the arts, not the sciences. For a time he was an apprentice in a painter's studio, and he later produced several canvases of his own, which were well received. He wrote poetry of some distinction and collaborated on an opera with his old school friend Germain Dandelin. Only through the influence of Jean Guillaume Garnier, a professor of mathematics at the newly created University of Ghent, was Quetelet finally persuaded to turn from his artistic endeavors to the full-time study of mathematics (although he continued to dabble in poetry until about the age of thirty). His doctoral

dissertation, in which he announced the discovery of a new curve, *la focale*, was the first to be presented at the university (on July 24, 1819), and was widely acclaimed as an original contribution to analytic geometry. As a result, at age 23 Quetelet found himself called to Brussels to occupy the chair of elementary mathematics at the Athenaeum. Only a few months later, early in 1820, he was further honored by being elected to membership in the Académie Royale des Sciences et des Belles-Lettres de Bruxelles.

The rapid pace set in the first part of Quetelet's scientific career never slackened; his productivity over the next fifty years was phenomenal. The record of his activities during the ten or so years after his arrival at Brussels well illustrates his prodigious capacity for work: Immediately after he assumed the position at the Athenaeum, he began publishing a vast array of essays, mostly in mathematics and in physics; they appeared initially in the *Nouveaux mémoires* of the academy and later in *Correspondances mathématiques et physiques*, a journal he founded in 1825 and edited (for the first two years together with Garnier) until its dissolution in 1839. For a time this was the foremost journal of its kind in Europe, attracting contributions from the most eminent scientists on the Continent. In 1824 Quetelet added to his duties the task of delivering a series of public lectures at the Brussels Museum, first in geometry, probabilities, physics, and astronomy and later in the history of the sciences. There, as at the Athenaeum, he quickly became renowned as a great teacher, and his lectures were always crowded, by regular students and auditors and by eminent scientists who came from all over Europe to hear him. He continued these lectures until 1834, when the museum was absorbed by the University of Brussels. Offered a chair in mathematics at the new university, he declined in order to devote himself to his many researches. However, the public lectures were resumed in 1836, at the Military School at Brussels, founded two years earlier.

During this same period Quetelet published several elementary works in natural science and in mathematics, designed to expose these fields to a wide popular audience. *Astronomie élémentaire*, published in 1826, was soon followed by *Astronomie populaire* (1827a). (Several biographers—notably Reichenberg [1896] and Hankins [1908]—asserted that the latter work immediately achieved the "distinction," accorded some earlier publications in astronomy, of being placed on the *Index librorum prohibitorum*. Lotin [1912, pp. 34–37] proved this to be a myth.) In 1827 Quetelet also published a



summary of his course in physics at the museum, entitled *Positions de physique, ou résumé d'un cours de physique générale* (1827b). Robert Wallace, in the preface to his English translation of the work in 1835, signaled its importance as follows: "No other work in the English language contains such an extensive and succinct account of the different branches of physics, or exhibits such a general knowledge of the whole field in so small a compass."

The following year, 1828, saw the publication of *Instructions populaires sur le calcul des probabilités*, which Quetelet identified as a résumé of the introductory lectures to his courses in physics and astronomy at the museum. This work marked Quetelet's shift from exclusive concentration on mathematics and the natural sciences to the study of statistics and, eventually, to the investigation of social phenomena.

Interestingly enough, Quetelet considered none of these works his main preoccupation or his main accomplishment in the ten years from 1823 to 1832. What concerned him most during this time was the project of establishing an observatory at Brussels. It is still unknown how Quetelet came to adopt this as a prime objective. What is certain, however, is that his activities directed toward establishing the observatory brought about, quite fortuitously, the major change of orientation in his scientific career. Upon accepting Quetelet's proposal for an observatory, the minister of education promptly sent him to Paris to acquaint himself with the latest astronomical techniques and instruments. There he was warmly received by the astronomers François Arago and Alexis Bouvard and was introduced by Bouvard to the coterie of French intellectuals gathered around the illustrious mathematicians Poisson, Laplace, and Jean Baptiste Fourier. These men had for some time been engaged in laying the foundations of modern probability theory, and several of them had analyzed empirical social data in their work.

It was instruction from these mathematicians, particularly Laplace, together with the stimulation of continual informal contact with their group at the École Polytechnique, that aroused in Quetelet the keen interest in statistical research and theory, based on the theory of probabilities, that was to become the focus of all his scientific work. In later reminiscences he said that after he had become acquainted with the statistical ideas of his French masters, he immediately thought of applying them to the measurement of the human body, a topic he had become curious about when he was a painter. One direct effect of learning the theory of probabilities was to make Quetelet realize "the

need to join to the study of celestial phenomena the study of terrestrial phenomena, which had not been possible until now. . . ." The crucial impact of the Paris experience on his thinking is evident a few sentences later, where he said, "Thus, it was among the learned statisticians and economists of that time that I began my labors . . ." (1870).

After his return from Paris in 1823, the project of the observatory moved along by fits and starts, at first held up by difficulties over financing and by some disagreements between Quetelet and the architect and later interrupted by the Belgian revolution of 1830 (during which the half-finished observatory was used as a makeshift fortress and suffered some structural damage). Quetelet finally took up residence in the nearly completed observatory in 1832.

### Organization of data collection

While the observatory was under construction, Quetelet's interest in statistics, which had crystallized during his visit to Paris, coupled with his manifest abilities as an organizer led him to become more and more active in projects requiring the collection of empirical social data. When the Royal Statistics Commission was formed in 1826, he became correspondent for Brabant. (From 1814 to 1830 Belgium was under Dutch rule; thus, "Royal" in this case refers to the House of Orange.) Quetelet's first publications covered quantitative information about Belgium which could be used for practical purposes, including mortality tables with special reference to actuarial problems of insurance. In 1827 he analyzed crime statistics, again with a practical eye to improving the administration of justice. In 1828 he edited a general statistical handbook on Belgium, which included a great deal of comparative material obtained from colleagues he had come to know during his stays in France and also in England. At his urging, a census of the population was taken in 1829, the results of which were published separately for Holland and Belgium after the revolution of 1830.

In 1841, largely through Quetelet's efforts, the Commission Centrale de Statistique was organized, and this soon became the central agency for the collection of statistics in Belgium. Quetelet served as its president until his death, and under his direction it performed its functions with remarkable thoroughness and efficiency, setting a standard for similar organizations throughout Europe. In 1833 he was delegated official representative to the meeting of the British Association for the Advancement of Science and there played a key role in the forma-

tion of a statistical section. Dissatisfied with the narrow scope of the section, he urged its chairman, Babbage, to organize the Statistical Society of London. This was accomplished in 1834, and the society survives today as the Royal Statistical Society (having been renamed in 1877).

In his work in statistics, as in the natural sciences, Quetelet placed great emphasis on the need for uniformity in methods of data collection and tabulation and in the presentation of results. His principal goal, in all his organizational endeavors, was to see this realized in practice. In 1851 Quetelet proposed to a group of scientists gathered at the Universal Exposition in London a plan for international cooperation in the collection of statistical information. The idea was heartily approved, and progress on it was so rapid that in 1853 the first International Statistical Congress was held at Brussels. At the initial session of the congress Quetelet was chosen president, and he naturally devoted his opening address to the importance of uniform procedures and terminology in official statistical publications. During the next twenty-five years the congress was enormously effective in spurring the development of official statistics in Europe, establishing permanent lines of communication between statisticians, and improving comparability. Internal dissension and controversy weakened the congress during the 1870s and eventually led to its collapse in 1880. By that time, however, Quetelet's original proposition that there must be some international organization to maintain uniformity and promote cooperation in the collection and analysis of official statistics had been so fully accepted that it was only a matter of five years before a new organization, the International Statistical Institute, was established to continue the work of the congress.

To round out this picture of Quetelet's successes as an organizer, we need only mention the pre-eminent role he played in the Académie Royale des Sciences et des Belles-Lettres. When he was chosen a member in 1820, the academy was near to closing down, with only about half a dozen superannuated members attending its sessions and virtually no publications to its name. Quetelet brought new life and vigor into the association, quickly assuming the major responsibility for its activities, recruiting into its ranks many of his young scientific colleagues, and fortifying its publications with his own numerous scientific writings. He was made director for the years 1832 and 1833, and in 1834 he was elected perpetual secretary, an office he occupied for the next forty years, during which time he was considered "the guiding spirit of the academy."

## Social research and theory

The two memoirs which form the basis for all of Quetelet's subsequent investigations of social phenomena appeared in 1831. By then he had decided that he wanted to isolate, from the general pool of statistical data, a special set dealing with human beings. He first published a memoir entitled *Recherches sur la loi de la croissance de l'homme* (1831a), which utilized a large number of measurements of people's physical dimensions. A few months later he published statistics on crime, under the title *Recherches sur le penchant au crime aux différents âges* (1831b). While the emphasis in these publications is on what we would call the life cycle, both of them also include many multivariate tabulations, such as differences in the age-specific crime rates for men and women separately, for various countries, and for different social groups. (As noted by Hankins [1908, p. 55] and by Lottin [1912, pp. 128-138], the core idea contained in the second memoir—the constancy in the "budget" of crimes from year to year in each age group—can be traced back to a memoir read to the academy on December 6, 1828, and published early in 1829. Thus, Quetelet was probably right in claiming priority for the idea [1835; see p. 96 in 1842 edition] over A. M. Guerry, who published it under the title "Statistique comparée de l'état de l'instruction et du nombre des crimes" [1832].)

In 1833 Quetelet published a third memoir giving developmental data on weight (1833a). By this time he had formed the idea of a social physics, and in 1835 he combined his earlier memoirs into a book entitled *Sur l'homme et le développement de ses facultés*, with the subtitle *Physique sociale* (see 1835). Quetelet republished this work in an augmented version, with the titles reversed, in 1869. To prevent confusion, the first edition is usually referred to as *Sur l'homme*, the second as *Physique sociale*. Included in the later edition, and one of its highlights, is a long essay by the English astronomer John Herschel: it had first been published in the *Edinburgh Review* (1850) as a review article on Quetelet's *Letters . . . on the Theory of Probabilities* and other work by Quetelet, and since it was highly favorable, Quetelet made it the introduction to his *Physique sociale*. To the constancy in crime rates noted earlier, this work added demonstrations of regularities in the number of suicides from year to year and in the rate of marriage for each sex and age cohort. Although Quetelet was convinced that many other regularities existed, these three—in crimes, suicides, and rates of marriage—were the only regularities in man's "moral" characteristics (i.e., those involv-



ing a choice of action) actually demonstrated in his writings. Thus, with the publication of *Sur l'homme*, all of Quetelet's basic ideas became available to a broader public.

**Basic principles.** It was in writings published in the 1830s that Quetelet established the theoretical foundations of his work in moral statistics or, to use the modern term, sociology. First there was the idea that social phenomena in general are extremely regular and that the empirical regularities can be discovered through the application of statistical techniques. Furthermore, these regularities have causes: Quetelet considered his averages to be "of the order of physical facts," thus establishing the link between physical laws and social laws. But rather than attach a theological interpretation to these regularities—as Süssmilch and others had done a century earlier, finding in them evidence of a divine order—Quetelet attributed them to social conditions at different times and in different places [see the biography of SÜSSMILCH]. This conclusion had two consequences: It gave rise to a large number of ethical problems, casting doubt on man's free will and thus, for example, on individual responsibility for crime; and in practical terms it provided a basis for arguing that meliorative legislation can alter social conditions so as to lower crime rates or rates of suicide.

On the methodological side, two key principles were set forth very early in Quetelet's work. The first states that "Causes are proportional to the effects produced by them" (1831*b*, p. 7). This is easy to accept when it comes to man's physical characteristics; it is the assumption that allows us to conclude, for example, that one man is "twice as strong" as another (the cause) simply because we observe that he can lift an object that is twice as heavy (the effect). Quetelet proposed that a scientific study of man's moral and intellectual qualities is possible only if this principle can be applied to them as well. (The role this principle played in Quetelet's theories is discussed below.) The second key principle advanced by Quetelet is that large numbers are necessary in order to reach any reliable conclusions—an idea that can be traced to the influence of Laplace (1812), Fourier (1826), and Poisson (1837). The interweaving of these principles with the theoretical ideas summarized above is illustrated in the following:

It seems to me that that which relates to the human species, considered en masse, is of the order of physical facts; the greater the number of individuals, the more the influence of the individual will is effaced, being replaced by the series of general facts that depend on the general causes according to which society

exists and maintains itself. These are the causes we seek to grasp, and when we do know them, we shall be able to ascertain their effects in social matters, just as we ascertain effects from causes in the physical sciences. (1831*b*, pp. 80–81)

Quetelet was greatly concerned that the methods he adopted for studying man in all his aspects be as "scientific" as those used in any of the physical sciences. His solution to this problem was to develop a methodology that would allow full application of the theory of probabilities. For in striking contrast to his contemporary Auguste Comte, Quetelet believed that the use of mathematics is not only the *sine qua non* of any exact science but the measure of its worth. "The more advanced the sciences have become," he said, "the more they have tended to enter the domain of mathematics, which is a sort of center toward which they converge. We can judge of the perfection to which a science has come by the facility, more or less great, with which it may be approached by calculation" (1828, p. 230).

**Pattern of work.** Before proceeding to a more detailed exposition of Quetelet's work in moral statistics, we should note his method of publication. Quetelet's literary background and the fact that his humanist friends remained an important reference group for him help to explain the manner in which he published his works. When he had new data or had developed a new technique or idea, he first announced his discovery in brief notes, usually in the reports of the academy or in *Correspondances* and sometimes in French or English journals. Once such notes had appeared, he would elaborate the same material into longer articles and give his data social and philosophical interpretations. He would finally combine these articles into books which he hoped would have a general appeal. He obviously felt very strongly that empirical findings should be interpreted as much as possible and made interesting to readers with broad social and humanistic concerns.

Quetelet further extended his influence through the voluminous correspondence he maintained with scientists, statesmen, and men of letters throughout Europe and America. Liliane Wellens-De Donder has identified approximately 2,500 correspondents, including such names as Gauss; Ampère; Faraday; Alexander von Humboldt; James A. Garfield, then a U.S. congressman, who solicited Quetelet's advice on means of improving the census; Joseph Henry; Lemuel Shattuck; Charles Wheatstone; Louis René Villermé; and Goethe, who befriended Quetelet when the latter visited Germany in 1829 (see Wellens-De Donder 1964).

Probably Quetelet's most famous correspondence was with the princes Ernest and Albert of Saxe-Coburg and Gotha, whom he tutored in mathematics beginning in 1836, at the request of their uncle, Leopold I, king of the Belgians. Although the princes left Belgium to attend school in Germany shortly after they began studying under Quetelet, the lessons continued for many years by correspondence. Quetelet's second major work on moral statistics, *Letters Addressed to H.R.H. the Grand Duke of Saxe-Coburg and Gotha, on the Theory of Probabilities, as Applied to the Moral and Political Sciences* (1846), shows his side of the correspondence. (The title refers only to Ernest, who as reigning duke was head of the house of Coburg.) *Du système social et des lois qui le régissent* (1848), was dedicated to Albert, with whom Quetelet had established an especially close friendship. Quetelet's profound influence on Albert's thinking is clearly shown in the keynote address Albert delivered to the fourth meeting of the International Statistical Congress in London, on July 16, 1860 (see Schoen 1938).

**The average man.** Quetelet's conceptualization of social reality is dominated by his notion of the average man, or *homme moyen*. In his preface to *Du système social*, he himself identified this as his central concept and traced its development through his writings (1848, pp. vii-ix). In *Sur l'homme*, he said, he had developed the idea that the characteristics of the average man can be presented only by giving the mean and the upper and lower limits of variation from that mean. In the *Lettres* he had shown that "regarding the height of men of one nation, the individual values group themselves symmetrically around the mean according to . . . the law of accidental causes" (p. viii); and further, that for a nation the average man "is actually the *type* or the standard and that other men differ from him, by more or by less, only through the influence of accidental causes, whose effects become calculable when the number of trials is sufficiently large . . . In this new work," Quetelet continued, "I show that the law of accidental causes is a general law, which applies to individuals as well as to peoples and which governs our moral and intellectual qualities just as it does our physical qualities. Thus, what is regarded as accidental ceases to be so when observations are extended to a considerable number of cases" (p. ix). It is no wonder, then, that discussions of Quetelet's theories and researches on society invariably take as a starting point his concept of the *homme moyen*.

His first approach to the concept was through

the measurement of physical characteristics of man, in particular height and weight (1831a). He conceived of the average height of a group of individuals of like age as the mean around which the heights of all persons of that age "oscillate," although just how this oscillation takes place Quetelet could not say. He did suggest, even in this first exposition of the concept, that similar means and oscillations might be observed if moral and intellectual, not just physical, qualities of men were studied.

In his essay *Recherches sur le penchant au crime*, the term *homme moyen* appears for the first time. There, also, we find the first statement of the idea that if one were to determine the *homme moyen* for a nation, he would represent the *type* for that nation; and if he were determined for all mankind, he would represent the *type* for the entire human species.

The next advance in the development of the concept came in a memoir published in 1844, in which Quetelet first took note of the fact that his observations were symmetrically distributed about the mean—in almost exactly the pattern to be anticipated (*prévu*) from the binomial and normal distributions—and went on to speculate about the likelihood that all physical characteristics might be distributed in the same way. By applying the theory of probabilities, he was then able to derive a theoretical frequency distribution for height, weight, or chest circumference that coincided remarkably with the empirical distributions in his data for various groups.

An interesting result obtained by applying this method, also first published in the memoir of 1844, was Quetelet's discovery of draft evasion in the French army. By noting the discrepancy between the distribution of height of 100,000 French conscripts and his prediction (i.e., the theoretical distribution calculated by assuming a probable error of 49 millimeters), he came to the conclusion that some 2,000 men had escaped service by somehow shortening themselves to just below the minimum height. Thus, quite by accident, Quetelet emerged with the first practical, although perhaps somewhat trivial, application of his statistical techniques.

In his discussions of the average man, Quetelet had up to this time limited himself to calculating the means and distributions of only a few physical characteristics. The task he now set for himself was to extend the concept to all of man's physical traits (thus forming the basis for what he called "social physics") and, thence, to all moral and intellectual qualities as well ("moral statistics").



Furthermore, he planned to apply the concept to collectivities of all sizes, ranging from the small group to the whole of mankind, and expected that it would hold equally well for any time in human history. Quetelet had suggested these extensions in earlier works, but the grand generalization did not emerge in its final form until the publication of *Du système social*. In the first pages Quetelet announced his theme: "There is a general law which governs our universe . . . ; it gives to everything that breathes an infinite variety. . . . That law, which science has long misunderstood and which has until now remained useless in practice, I shall call the *law of accidental causes*" (1848, p. 16). A few lines later he elaborated his over-all viewpoint:

. . . among organized beings all elements vary around a mean state, and . . . variations, which arise from accidental causes, are regulated with such harmony and precision that we can classify them in advance numerically and by order of magnitude, within their limits.

One part of the present work is devoted to demonstrating the law of accidental causes, both for physical man and for moral and intellectual man, considering him individually, as well as in the aggregate. . . . (p. 17)

**Concept of causality.** It is obvious that an explanation of what Quetelet meant by "accidental causes" and by "law" is critical for an understanding of his conception of the average man. Quetelet hypothesized that every mean he presented resulted from the operation of constant causes, while the variations about the mean were due to "perturbative" or "accidental" causes. "Constant causes," he explained, "are those which act in a continuous manner, with the same intensity and in the same direction" ([1846] 1849, p. 107). Among the constant causes he named are sex, age, profession, geographical latitude, and economic and religious institutions. (As a category parallel to constant causes, Quetelet sometimes mentioned "variable causes," which are those that "act in a continuous manner, with energies and intensities that change" [*ibid.*]. The seasons are cited as the type case, although Quetelet meant to include as variable causes all periodical phenomena.) "Accidental causes only manifest themselves fortuitously, and act indifferently in any direction" (*ibid.*). Quetelet frequently classed man's free will as an accidental cause (although occasionally he claimed that it played no role at all), but insisted that its operation is constrained within very narrow limits. The essence of Quetelet's theory is that,

given sufficient data over time, the shape and extent of variations about the mean state which result from accidental causes can be "classified in advance" with a high degree of accuracy, through the application of the theory of probabilities of independent events.

Quetelet's conception of "law" depended on whether he was talking about man's physical attributes, his moral traits, or all human characteristics. Thus, in *Du système social* we find these three distinct uses of the term. In the early part of the book, he referred to a trend in a series of averages over time as a law: "If we knew what [man's mean] height had been from one century to another, we would have a series of sizes which would express the law of development of humanity as regards height" (1848, p. 11). Later, in presenting the law of propensity to crime, he used the term to denote a regular pattern of correlations: "*The propensity to crime increases quite rapidly toward adulthood; it reaches a maximum and then decreases until the very end of life.* This law appears to be constant, and varies only with respect to the magnitude of the maximum and the time of life when it occurs" (p. 86). (By way of contrast, his law of propensity to suicide posits a direct variation with age "until the most advanced age" [p. 88].)

It is important to note in this connection that Quetelet went far beyond such simple two-variable correlations in his studies of man. Numerous three-variable and four-variable tables appear throughout his work (see esp. 1835, vol. 2). In one case, for example, he presented a table that shows the relationship of mean weight to age, sex, and occupation (*ibid.*, p. 91); similar tables show the breakdown of crimes in various groups by sex and level of education (p. 297), by age and sex (p. 302), and by age, sex, and the type of court in which the crime was tried (p. 308). These remarkable anticipations of modern techniques went largely unnoticed by Quetelet's contemporaries, and only in recent times have social scientists rediscovered and fully explored the possibilities of multivariate analysis—a striking discontinuity in the history of empirical social research that surely deserves further study and explanation.

The two uses of the term "law" illustrated above are similar in that they both refer in some way to a correlation—in one case between height and century, and in the other between age and the incidence of some social act. Quetelet's third type of law, the "law of accidental causes," is quite another thing; it is simply the assertion that every human trait is normally distributed about a mean

and that the larger the number of observations, the more closely the empirical distribution will coincide with the theoretical probability distribution. In sum, the word "law" is used alternatively to refer to a trend in a series of specific empirical findings, an empirical generalization, and an assertion (or, in effect, a theory) that a certain type of regularity exists in all human phenomena.

**Measurement.** Perhaps one reason Quetelet had trouble maintaining a single conception of "law" is that the types of measures he used to substantiate laws were few in number and inadequate to his purpose. Limiting himself to the manipulation of data gathered from available official statistical publications, he was forced to improvise new and different techniques for physical and moral qualities and so emerged with different kinds of laws, according to the type of phenomenon in question. It comes as something of a surprise to realize that in all Quetelet's research on man, for example, he actually used only three kinds of measurements. (He did suggest some others but never applied them to his data or attempted to collect the data that would make them applicable. Lazarsfeld's 1961 essay analyzes the kinds of measures that are mentioned in *Physique sociale* in the light of modern ideas on quantification.)

First, he determined the empirical distribution of some human trait in a group and computed the mean—which he then identified as a characteristic of the average man for the group. He repeatedly asserted that similar distributions and means could be found for moral and intellectual characteristics, presented some rough hypothetical curves, but never performed any such calculations using a set of empirical observations. A second type of measurement involved counting the number of certain social events, such as crimes or marriages, that occurred during a series of years among particular groups; the average of these yearly counts was taken to be the probable number of such events that would occur in each group during the next year. As a third type of measurement Quetelet used rates—the number of crimes or other events in each age group, divided by the number of persons in that group. Quetelet regarded the results as the respective probabilities of committing a crime at various ages; he called these probabilities the propensity to crime—the *penchant au crime* or, alternatively, the *tendance au crime*—at each age. (At one point in *Du système social* [1848, p. 93] he proposed substituting the word *possibilité* for *penchant*, but he reverted to the use of *penchant*

throughout the later text.) Although Quetelet said again and again that his concern was with groups, not with individuals, there are many instances where he clearly uses the *penchant au crime* derived by this method to refer to a characteristic of *each* member of a given age group.

**Apparent and real propensities.** Quetelet developed the concept of *penchant* in order to overcome the methodological and theoretical problems he encountered in trying to found a science that would deal with all aspects of man. As long as he restricted himself to physical characteristics there was no problem; the technique for obtaining individual measurements was obvious, and moreover, plentiful data of this sort had already been collected by many agencies and was available for analysis. Once he moved on to moral and intellectual traits, however, the only data available were rates (of crime, or suicide, or marriage) for different populations. To parallel his analysis of physical characteristics, Quetelet would have needed measurements taken on each individual in a group over a period of time, and so far no one had collected such data. Occasionally, in the more speculative portions of his writings, Quetelet was able to suggest how such individual measurements might be made: to measure a scholar's productivity, for example, he thought of counting the number of publications the man produced.

Unfortunately, he never applied such ideas in his statistical work. Instead, he made do with the data at his disposal by establishing the critical distinction between "apparent propensities" and "real propensities." Apparent propensities are those that can be calculated as outlined above, using the population rates found in official statistical publications. The information needed is only the number of acts (crimes, suicides, marriages), the age of each actor, and the total population, distributed according to age. The real propensity is what causes the observable regularity to appear; and this propensity, Quetelet claimed, cannot be ascertained from direct observation. It can be known only by its effects.

The following passage illustrates how Quetelet related the two types of propensities. Commenting on computations of the probability of marriage for certain city dwellers, he said:

This probability may be considered as giving, in cities, the measure of the *apparent tendency* to marriage of a Belgian aged 25 to 30. I say *apparent tendency* intentionally, to avoid confusion with the *real tendency*, which may be quite different. One man may have, throughout his life, a real tendency to marry without



ever marrying; another, on the contrary, carried along by fortuitous circumstances, may marry without having the least propensity to marriage. The distinction is essential. (1848, p. 77)

The cue to the way Quetelet visualized the relation between apparent tendencies and real tendencies lies in his repeated statement that "causes are proportional to effects." Thus, if one thinks of the apparent tendencies as being caused by the underlying real tendencies, "the error that may result from substituting the value of the one for the other can be calculated directly by the theory of probabilities" (*ibid.*, p. 78).

Rather than solve the problem, this does no more than identify it. Quetelet's difficulty arose from the fact that he never clearly separated the problem into its two components. One is the question that continues to be of interest today: how may manifest data be related to latent dispositions? The other is the methodological problem of whether measurement techniques analogous to those applied to individual physical traits could be developed for moral and intellectual characteristics as well.

On the theoretical side, Quetelet failed to recognize that his dispositional concept of *penchant* could just as reasonably be applied in the study of physical attributes as in the study of moral or intellectual traits. One can, for example, conceive of studying the "tendency to obesity" in a population, which would parallel, in the physical, Quetelet's notion of a *penchant au crime*. On the methodological side, he failed to realize that some of the techniques he himself suggested for measuring individual personality or intelligence were exact counterparts to his "direct" measurements of physical traits. Especially surprising—in view of his extensive use of crime statistics—is the fact that the idea of analyzing, for example, "repeated offenders," completely eluded him. This would have provided him with quantitative individual "measures" of criminal behavior, corresponding to measures of size or weight. It seems likely that these shortcomings in Quetelet's work were due, not to sheer lack of insight, but, at least in part, to the inadequacies of the data available at the time. It is only benefit of hindsight that allows us to identify such gaps, and one cannot gainsay Quetelet's merit in having made the first attempt to handle what remain to this day crucial problems in the analysis of empirical social phenomena.

#### After Quetelet

In 1855 Quetelet suffered a stroke, from which he never totally recovered. He resumed his work

very soon afterward but never again produced any new ideas. His publications from then on, although numerous, were largely compendia of prior essays or summaries of new researches which supported his earlier ideas. His son Ernest virtually took over the running of the observatory after 1855. Quetelet died on February 17, 1874, and as Hankins put it, "was buried with honors befitting one of the earth's nobility." A statue of him, funded by popular subscription, was unveiled at Brussels in 1880.

Quetelet's concern with the distribution of human characteristics was destined for an interesting future. His basic idea was that certain social processes (corresponding to his interplay of causes) would explain the final distribution of certain observable data. This notion has been amply justified by modern mathematical models regarding the distribution of, for example, income, words, or city sizes. But Quetelet concentrated exclusively on the binomial and normal (Gaussian) distributions, which presuppose the independence of the events studied. Today many other distributions are known, based on more complicated processes; especially in the social sciences, "contagious events," which depend upon each other, are in the center of interest. Quetelet remained unaware of alternative mathematical possibilities. Nonetheless, his basic idea was not only correct but probably influenced directly writers who had begun to broaden this whole field, such as Poisson and Lexis.

In general, however, Quetelet's contemporaries focused their attention primarily on his concept of the average man and his proposition that social phenomena reproduce themselves with extreme regularity. For different reasons these ideas quickly became the subject of the most vigorous and widespread debate among nineteenth-century statisticians, philosophers, and social scientists, while most of Quetelet's other ideas remained largely unnoticed.

The debate over the *homme moyen* was set off by Quetelet's suggestion that the means of various traits could be combined to form one paradigmatic human being, who would represent the "type" for a group, a city, a nation, or even for all of mankind. Typical of the early criticism was that of Cournot (1843), who reasoned from a mathematical analogy: just as the averages of the sides of many right triangles do not form a right triangle, so the averages of physical traits would certainly not be compatible. Combining them would not, as Quetelet claimed, produce a "type for human beauty" or a "type for physical perfection," but a monstrosity. Quetelet's insistence that the average

man be considered no more than a "fictitious being" was taken as simply an evasion of the issue; his attempts to reply directly to Cournot proved unconvincing. (A recent essay by Guilbaud discusses the Cournot problem as it relates to current statistical concepts like that of "aggregation"; Guilbaud 1952.)

Quetelet's inability to refute Cournot's criticism encouraged others to publish similar attacks until, in 1876, Bertillon issued what is usually considered the definitive statement, which pretty well put an end to the debate. Here the criticism was applied not only to Quetelet's notion of combining average physical characteristics but also to his ideas about moral and intellectual traits. Surely, Bertillon said, Quetelet was mistaken in believing that his average man would represent the ideal of moral virtue or intellectual perfection. Such a man would, on the contrary, be the personification of mediocrity or, to use Bertillon's apt phrase, the *type de la vulgarité* (p. 311). Thereafter the average man was generally viewed as a concept not worth taking seriously, although sporadic attempts have been made to revive it. (The latest is that of Maurice Fréchet, who suggests that Quetelet's *homme moyen* could be "rehabilitated" as the concept of the *homme typique*; by defining the *homme typique* as a particular individual in the group—whose traits taken as a whole come closest to the average—Fréchet manages to avoid most of the criticisms leveled against Quetelet's concept; Fréchet 1955.)

A second and even more widespread controversy centered on the question of what implications ought to be drawn from the startling regularities Quetelet had demonstrated in his studies of social phenomena. Did Quetelet's proposition that "society prepares the crime and the guilty person is only the instrument by which it is executed" (1835, p. 108 in 1842 edition) imply that human beings have no free will at all? Quetelet's philosophical speculations on the subject certainly left room for this interpretation. The result was a heated and long-lasting debate between those who supported Quetelet's "deterministic" explanation of social regularities and those who argued that only by taking the individual as the starting point for analysis could one arrive at an explanation of human behavior. Free will, the latter group contended, must be considered a prime determinant of action, not classed as a practically negligible "accidental cause." (One by-product of the debate was the formation of a "German school" of moral statisticians, headed by Moritz Wilhelm Drobisch, one of Quetelet's most vehement opponents; see

Drobisch 1867.) The over-all result of the controversy was not so much to refute Quetelet's ideas as to argue them into oblivion. In P. E. Fahlbeck's opinion (1900), its most important effect was that until the end of the nineteenth century statisticians were so involved in discussing the implications of Quetelet's propositions that they made little effort to confirm empirically the nature and extent of the regularities Quetelet had discovered. (A detailed discussion of the controversy over Quetelet's determinism and over the concept of the average man appears in Lottin 1912, pp. 413-458.)

Only in recent years has Quetelet's sociological work begun to receive due recognition. His conviction that a scientific study of social life must be based on the application of quantitative methods and mathematical techniques anticipated what has become the guiding principle of modern social research. Some of the specific methods he employed and advocated—e.g., the substitution of one-time observation of a population for repeated observations of the individual, and his early attempts at multivariate analysis—are as important today as they were new in his time. The same may be said of his efforts to transform statistics from the mere clerical task of collecting important facts about the state (hence the term "statistics") to an exact method of observation, measurement, tabulation, and comparison of results, which would serve as the scaffolding upon which he could erect his science of moral statistics. On these grounds alone it is difficult to dispute Sarton's description of *Sur l'homme* as "one of the greatest books of the nineteenth century" ([1935] 1962, p. 229); or, for that matter, his choice of Quetelet over Comte as the "founder of sociology."

DAVID LANDAU AND PAUL F. LAZARSFELD

[For the historical context of Quetelet's work, see SOCIOLOGY, article on THE EARLY HISTORY OF SOCIAL RESEARCH, and the biographies of GAUSS; LAPLACE; POISSON; for discussion of the subsequent development of Quetelet's ideas, see GOVERNMENT STATISTICS; SOCIOLOGY, article on THE FIELD; STATISTICS, DESCRIPTIVE, article on LOCATION AND DISPERSION; and the biographies of BERTILLON; GALTON; GINI; NICEFORO; PEARSON.]

#### WORKS BY QUETELET

- (1826) 1834 *Astronomie élémentaire*. 3d ed., rev. & corrected. Brussels: Tircher.
  - (1827a) 1832 *Astronomie populaire*. 2d ed., rev. Brussels: Remy.
  - (1827b) 1834 *Positions de physique, ou résumé d'un cours de physique générale*. 2d ed. Brussels: Tircher.
- An English translation was published by Sinclair



- in 1835, as *Facts, Laws and Phenomena of Natural Philosophy: Or, Summary of a Course of General Physics*.
- 1828 *Instructions populaires sur le calcul des probabilités*. Brussels: Tarlier. → Translation of extract in the text was provided by David Landau. An English translation of the entire work was published by Weale in 1849, as *Popular Instructions on the Calculation of Probabilities*, with appended notes by Richard Beamish.
- 1831a *Recherches sur la loi de la croissance de l'homme*. Brussels: Hayez. → A 32-page pamphlet. Also published in Volume 7 of the *Nouveaux mémoires* of the Académie Royale des Sciences, des Lettres et des Beaux-Arts de Belgique.
- (1831b) 1833 *Recherches sur le penchant au crime aux différents âges*. 2d ed. Brussels: Hayez. → An 87-page pamphlet. Translations of extracts in the text were provided by David Landau. Also published in Volume 7 of the *Nouveaux mémoires* of the Académie Royale des Sciences, des Lettres et des Beaux-Arts de Belgique.
- 1833a *Recherches sur le poids de l'homme aux différents âges*. Brussels: Hayez. → A 44-page pamphlet. Also published as part of Volume 7 of the *Nouveaux mémoires* of the Académie Royale des Sciences, des Lettres et des Beaux-Arts de Belgique.
- 1833b Lettre à M. Villermé, sur la possibilité de mesurer l'influence des causes qui modifient les éléments sociaux. *Annales d'hygiène publique et de médecine légale* 1st Series 9:309 only.
- (1835) 1869 *Physique sociale: Ou, essai sur le développement des facultés de l'homme*. 2 vols. Brussels: Muquardt. → First published as *Sur l'homme et le développement de ses facultés: Physique sociale*. An English translation was published by Chambers in 1842 as *A Treatise on Man and the Development of His Faculties*.
- 1844 *Sur l'appréciation des documents statistiques, et en particulier sur l'appréciation des moyennes*. Belgium, Commission Centrale de Statistique, *Bulletin* 2:205-286.
- (1846) 1849 *Letters Addressed to H.R.H. the Grand Duke of Saxe-Coburg and Gotha, on the Theory of Probabilities, as Applied to the Moral and Political Sciences*. London: Layton. → First published in French.
- 1848 *Du système social et des lois qui le régissent*. Paris: Guillaumin. → Translations of extracts in the text were provided by David Landau.
- 1870 *Des lois concernant le développement de l'homme*. Académie Royale des Sciences, des Lettres et des Beaux-Arts de Belgique, *Bulletin* 2d Series 29:669-680. → Translations of extracts in the text were provided by David Landau.
- Discours [prononcés aux funérailles de Quetelet], by N. de Keyser et al. 1874 Académie Royale des Sciences, des Lettres et des Beaux-Arts de Belgique, *Bulletin* 2d Series 37:248-266
- DROBISCH, MORITZ W. 1867 *Die moralische Statistik und die menschliche Willensfreiheit*. Leipzig: Voss.
- DURKHEIM, ÉMILE (1897) 1951 *Suicide: A Study in Sociology*. Glencoe, Ill.: Free Press. → First published in French. See pages 300-306 for a discussion of Quetelet's concept of the "average man."
- FAHLBECK, PONTUS E. 1900 La régularité dans les choses humaines ou les types statistiques et leurs variations. *Journal de la Société de Statistique de Paris* 41:188-201.
- [FOURIER, JEAN BAPTISTE] 1826 Mémoire sur les résultats moyens déduits d'un grand nombre d'observations. Volume 3, pages ix-xxxi in *Seine (Dept.), Recherches statistiques sur la ville de Paris et le département de la Seine*. Paris: Imprimerie Royale.
- FRÉCHET, MAURICE 1955 Rehabilitation de la notion statistique de l'homme moyen. Pages 310-341 in Maurice Fréchet, *Les mathématiques et le concret*. Paris: Presses Universitaires de France.
- GALTON, FRANCIS (1869) 1952 *Hereditary Genius: An Inquiry Into Its Laws and Consequences*. New York: Horizon Press. → See the Appendix for a discussion of some of the mathematical aspects of the "average man." A paperback edition was published in 1962 by World.
- GILLISPIE, C. C. 1963 Intellectual Factors in the Background of Analysis by Probabilities. Pages 431-453 in *Symposium on the History of Science*, Oxford, 1961, *Scientific Change: Historical Studies in the Intellectual, Social, and Technical Conditions for Scientific Discovery and Technical Invention, From Antiquity to the Present*. New York: Basic Books.
- GINI, CORRADO 1914a *L'uomo medio. Giornale degli economisti e rivista di statistica* 3d Series 48:1-24.
- GINI, CORRADO 1914b *Sull'utilità delle rappresentazioni grafiche. Giornale degli economisti e rivista di statistica* 3d Series 48:148-155.
- GUERRY, ANDRÉ MICHEL 1832 Statistique comparée de l'état de l'instruction et du nombre des crimes. *Revue encyclopédique* 55:414-424.
- GUILBAUD, GEORGES TH. (1952) 1966 Theories of the General Interest and the Logical Problem of Aggregation. Pages 262-307 in Paul F. Lazarsfeld and Neil W. Henry (editors), *Readings in Mathematical Social Science*. Chicago: Science Research Associates. → See especially pages 271-292. First published in French in Volume 5 of *Économie appliquée*.
- HALBWACHS, MAURICE 1912 *La théorie de l'homme moyen: Essai sur Quetelet et la statistique morale*. Paris: Alcan.
- HANKINS, FRANK H. 1908 *Adolphe Quetelet as Statistician*. New York: Longmans.
- [HERSCHEL, JOHN F. W.] 1850 [Review of] *Letters*. . . *Edinburgh Review* 92:1-57.
- KNAPP, GEORG F. 1871 Bericht über die Schriften Quetelet's zur Socialstatistik und Anthropologie. *Jahrbücher für Nationalökonomie und Statistik* 17:167-174, 342-358, 427-445.
- KNAPP, GEORG F. 1872 A. Quetelet als Theoretiker. *Jahrbücher für Nationalökonomie und Statistik* 18:89-124.
- LAPLACE, PIERRE SIMON DE (1812) 1820 *Théorie analytique des probabilités*. 3d ed., rev. Paris: Courcier.

## SUPPLEMENTARY BIBLIOGRAPHY

- BERTILLON, ADOLPHE 1876 La théorie des moyennes en statistique. *Journal de la Société de Statistique de Paris* 17:265-271, 286-308.
- BUCKLE, HENRY THOMAS (1857-1861) 1913 *The History of Civilization in England*. 2d ed. 2 vols. New York: Hearst.
- Correspondances mathématiques et physiques*. → Published from 1825 to 1839.
- COURNOT, ANTOINE AUGUSTIN 1843 *Exposition de la théorie des chances et des probabilités*. Paris: Hachette. → See especially pages 213-214.

- LAZARSFELD, PAUL F. 1961 Notes on the History of Quantification in Sociology: Trends, Sources and Problems. *Isis* 52:277-333.
- LOTTIN, JOSEPH 1912 *Quetelet, statisticien et sociologue*. Paris: Alcan; Louvain: Institut Supérieur de Philosophie.
- MAILLY, NICOLAS É. 1875a Essai sur la vie et les ouvrages de Lambert-Adolphe-Jacques Quetelet. *Académie Royale des Sciences, des Lettres et des Beaux-Arts de Belgique, Annuaire* 41:109-297.
- MAILLY, NICOLAS É. 1875b Eulogy on Quetelet. *Smithsonian Institution, Annual Report* [1874]:169-183. → Abstract of "Notice sur Adolphe Quetelet," first published in Series 2, Volume 38 of the *Bulletin of the Académie Royale des Sciences, des Lettres et des Beaux-Arts de Belgique*.
- POISSON, SIMÉON DENIS 1837 *Recherches sur la probabilité des jugements en matière criminelle et en matière civile, précédées des règles générales du calcul des probabilités*. Paris: Bachelier.
- REICHESBERG, NAUM 1896 Der berühmte Statistiker, Adolf Quetelet, sein Leben und sein Wirken: Eine biographische Skizze. *Zeitschrift für schweizerische Statistik* 32:418-460.
- SARTON, GEORGE (1935) 1962 Quetelet (1796-1874). Pages 229-242 in George Sarton, *Sarton on the History of Science*. Edited by Dorothy Stimson. Cambridge, Mass.: Harvard Univ. Press.
- SCHOEN, HARRIET H. 1938 Prince Albert and the Application of Statistics to the Problems of Government. *Osiris* 5:276-318.
- WELLENS-DE DONDER, LILIANE 1964 La correspondance d'Adolphe Quetelet. *Archives et bibliothèques de Belgique* 35:49-66.
- WOLOWSKI, L. 1874 Éloge de Quetelet. *Journal de la Société de Statistique de Paris* 15:118-126.

## QUEUES

There are many situations in industry and everyday life in which customers require service from limited service facilities. Much work has been done on techniques for predicting the amount of congestion in such systems. Applications include the following: telephone calls requiring service at an exchange; aircraft requiring an opportunity to take off or land; cars requiring a changed traffic light, or an opportunity to turn from a minor road into a major road; products at one stage of an industrial process requiring entrance to the next stage of processing; machines requiring attention from an operator; people requiring transport by bus, train, or airplane; and patients requiring attention by a doctor or admission to a hospital.

The objective of most investigations of queueing is to modify the system to make it more efficient. For example, the object of a study of queueing at traffic lights might be to formulate rules for adjusting the timing of the lights to the rates of flow of traffic. Such matters fall within the general subject

of operations research. There seems to be as yet little work on queueing of a fundamental sociological nature.

**Methods of investigation.** There are, essentially, four methods of investigating queueing problems: (1) direct observation of practical situations, (2) planned experiments under artificial conditions, (3) simulation, and (4) mathematical analysis. Published work is largely on mathematical analysis, but this does not properly reflect the relative importance of the methods. The last sections of this article deal with simulation and mathematical analysis.

**Some elementary results.** Some important theoretical ideas and results can be illustrated with a simple example. Consider customers arriving at a shop having one server who serves only one customer at a time. Unserved customers queue up waiting their turn for service. Suppose that the instants at which customers arrive are distributed in a stable statistical pattern in which the mean interval between the arrival of successive customers is  $a$  minutes. Let the time taken to serve a customer have a stable frequency distribution with mean  $s$  min.

In a very long time,  $T$  min., about  $T/a$  customers will arrive, whereas, even if the server works continuously, the number of customers served in that time is about  $T/s$ . Hence, if  $T/a > T/s$ , that is, if  $s/a > 1$ , customers arrive more rapidly than they can be served and the queue will grow until the system changes; for example, a second server may be obtained or customers may be deterred from joining the queue.

The first step in many queueing investigations is to see whether the average service capability is enough to meet average demands. Suppose that this is so, that is,  $s/a < 1$ . It is reasonable to expect a state of statistical equilibrium in which the number of customers queueing fluctuates with a stable frequency distribution. Also, in a long time  $T$ , the server will need to work for a total time of about  $sT/a$ , in order to serve  $T/a$  customers. Hence, there will be no customers in the system for a proportion  $1 - (s/a)$  of the time. The dimensionless quantity  $s/a$ , called the *traffic intensity*, measures the average demand on the system relative to the average service capability. The qualitative behavior of the system is determined in the first place by whether the traffic intensity exceeds, or is less than, one (Nearly always, the case of exact balance is unstable, leading eventually to very long queues.)

The next general point is that when  $s/a < 1$  the amount of congestion depends on the random fluctuation of arrivals.



tuations present. Thus, suppose that  $a = 1$ ,  $s = 0.9$ . If arrivals are perfectly regular every minute and if service always takes 0.9 min., an exact cycle of working is set up and no congestion arises. But if there is appreciable random variation in the arrivals or in the service times, there will, from time to time, be several customers at the service point simultaneously, and congestion will result. It can be shown that if arrivals are completely random (defined later) and the service time is constant, then the mean time spent queueing before the start of service is 4.05 min. and the server is idle 10 per cent of the time. Further, if both arrivals and service times are random and if the standard deviation of service time is equal to the mean, the mean time spent queueing rises to 8.1 min. With regular arrivals and a particular form of distribution of service time with standard deviation equal to the mean, the mean queueing time is 3.76 min. These represent substantial congestion.

The important general points are (1) if there is appreciable random variation, substantial congestion may result if the system is loaded near to its limiting average capacity, and (2) the amount of congestion depends in an essential way on the statistical variability in the system.

The amount of congestion can be reduced in various ways, for example, by having more servers, by reducing the variability in the system, or by reducing the traffic intensity. Suppose, however, that the traffic intensity is reduced to 0.8, by reduction of the service time to a constant,  $s = 0.8$  min. Then the mean queueing time drops from 4.05 min. to 1.6 min. The proportion of time that the server is idle rises from 0.1 to 0.2. An analysis of the costs associated with congestion and with the service mechanism is usually required to find the best way to run the system.

*General description of queueing systems.* The main features that determine a queueing system are the input or arrival system, the service mechanism, and the queue discipline. Many possibilities arise, only a few being described here.

The main features of the input are the average rate of arrival (including its dependence, if any, on time) and the local statistical character of the arrival pattern. The two simplest types of arrival used in mathematical work are *regular* arrivals and *random* arrivals. The latter, often called a Poisson process, is a very special form defined as follows. In any small time interval  $(t, t + \Delta t)$  there is a chance  $\Delta t/a$  of a customer arriving, quite independently of arrivals in other time periods. The average interval between arrivals is  $a$ ; the distribu-

tion of  $x$ , the interval between successive arrivals, is negative exponential, having density function  $a^{-1} \exp(-x/a)$ ,  $x > 0$  [see DISTRIBUTIONS, STATISTICAL].

The service mechanism is described by the service capacity, the service availability, and the properties of service time. The service capacity is the maximum number of customers that can be served at a time, one in the single-server queue discussed above. In some applications—for example, many transport problems—service is available only at certain times, and this must be specified; or service may not be available until a group of customers of specified size has collected (batch service). Finally, one must specify the statistical properties of the service time, usually by giving a probability distribution. Perhaps the most common type has a unimodal positively skewed density function with a standard deviation of 20 to 50 per cent of the mean. In mathematical work, a particularly important density is the negative exponential, with equation  $s^{-1}e^{-s/\tau}$ , where  $s$  is the mean service time.

The queue discipline specifies what happens to customers who cannot be served immediately. Sometimes there may be losses, for example, customers may have to be rejected because of lack of waiting room. Then the probability of loss should be calculated. If there are no losses, customers queue awaiting their turn for service. One then needs to specify how customers are selected for service from the queue. Even in a single-server system there are many possibilities: for example, first-come, first-served, that is, customers queue in order of arrival; random selection from a queue of customers; non-pre-emptive priority, in which the customers are divided into two or more priority classes so that when a customer is to be selected for service the one with the highest priority is chosen; and pre-emptive priority, in which a high-priority customer is served immediately upon arrival.

This classification applies to a single point of congestion. Often there will be several such inter-linked points.

Arrival pattern, service mechanism, and queue discipline determine the system. The resulting amount of congestion can be described by the distributions of three quantities: the number of customers in the system at an arbitrary time, the time a customer spends queueing before being served, and the server's busy and idle periods. Which quantity or combination of quantities should be considered depends on the costs associated with congestion.

**Investigating the behavior of queues.** Several of the usual methods used in investigating queues are described below.

*Simulation.* The structure of the system is specified statistically, for example, by giving numerically the frequency distribution of service time. The behavior of the system is then *simulated*, that is, reconstructed empirically using pencil and paper (in simple cases), an electrical or hydraulic analogue machine, or, most commonly, digital computation on an electronic computer. The behavior over a long period is found and relevant properties, such as queueing times, measured. The procedure is normally repeated for a range of values of the important parameters. This approach is empirical observation of an idealized statistical model. It is sometimes called the Monte Carlo method (Florida . . . 1956). [See COMPUTATION.]

Simulation is not necessary for the one-server queueing model for which numerical results were given above, because simple mathematical results can be found. If used, however, simulation would proceed roughly as follows. From tables of random numbers, or from a computer program for generating pseudo-random numbers, values are formed representing the intervals between the arrivals of customers and the service times of customers. The arrival instants and the instants at which service is started and completed can be built up and anything of interest, such as the frequency distribution of queueing time, measured [see SIMULATION].

An important advantage of this method is its extreme flexibility. Systems far too complex for mathematical analysis can be investigated. A disadvantage is that many simulations under a wide range of conditions may be necessary to understand fully the system's behavior. Simulation is, however, the most generally applicable technique for investigating specific practical problems.

*Mathematical analysis.* The theory of stochastic processes deals with systems that change in time probabilistically. To get working results about queues from the theory, it is usually necessary to consider very simplified models of the real problem. The usefulness of the theory is partly that skillful simplification may leave the essential features preserved and, perhaps more importantly, that intensive study of simplified models can lead to valuable qualitative understanding. A good example of qualitative conclusions arising from a simplified mathematical model comes from an investigation by Tanner (1961) of a model of overtaking on a two-lane road. Tanner's calculations show that under certain circumstances there is a density of slow

traffic, well below the capacity of the road, above which overtaking is effectively impossible. In many circumstances, high acceleration is more important than high speed, and a cutting of safety margins makes little difference to the average speed of a fast car over its whole journey. Mathematical investigation of simplified models combined with simulation of more complex models is a powerful method of investigation.

Essentially three types of mathematical investigation can be made: (1) to find rigorously the conditions for the existence of a unique equilibrium statistical distribution, (2) to find the equilibrium distribution assuming that it exists, and (3) to examine the transient behavior when, for example, the parameters change in time. The answer to (1) is usually qualitatively obvious, although the rigorous proofs often call for delicate arguments. Answers to (2) form the bulk of the literature and are relevant when long-run behavior is investigated. When behavior over a short period or under changing conditions is required, the transient behavior needs consideration, but not many usable results are yet available.

*Application of Markov processes.* The general idea behind the mathematical theory is to relate the probability distributions of the states of the system at different times. Consider the transition probabilities determining the probability distribution of the system at time  $t + \Delta t$ , given the state at time  $t$ . If these transition probabilities depend only on the state at  $t$  and not also on what happens before  $t$ , the process is a Markov process [see MARKOV CHAINS]. It is in principle then easy to find simultaneous differential equations describing the transient behavior of the system and ordinary linear equations for the equilibrium behavior.

The Markov property holds in simple form when arrivals are random and the distribution of service time is negative exponential, with density  $s^{-1}e^{-s/s}$ . Simplicity is gained by having random arrivals because the probability of an arrival in the small time interval  $(t, t + \Delta t)$  is  $\Delta t/a$ , *independently of all other occurrences*. The analogous property of the exponential distribution of service time is that if service of a customer is in progress at time  $t$ , the probability that service is completed in  $(t, t + \Delta t)$  is  $\Delta t/s$ , where  $s$  is the mean service time, *independently of the length of time for which the particular service operation has been going on*.

For example, consider the single-server queue with random arrivals and exponentially distributed service time. Let  $p_n(t)$  be the probability at time  $t$  that there are  $n$  customers in the system, including



the one, if any, being served. Consider  $p_n(t + \Delta t)$ , where  $\Delta t > 0$  is very small. The ways in which the queue may be empty at  $t + \Delta t$  are that (1) the queue is empty at  $t$  and no customer arrives in  $(t, t + \Delta t)$ , (2) there is one customer present at  $t$  whose service is completed in  $(t, t + \Delta t)$ , and (3) two or more events occur in  $(t, t + \Delta t)$ . The probability of (3) is negligible. Thus  $p_0(t, t + \Delta t)$  is the sum of the probabilities of (1) and (2). That is,

$$p_0(t + \Delta t) = p_0(t)\{1 - \Delta t/a\} + p_1(t)\Delta t/s,$$

from which

$$(1) \quad \frac{dp_0(t)}{dt} = -\frac{1}{a}p_0(t) + \frac{1}{s}p_1(t).$$

Similarly, if  $n \geq 1$ ,  $p_n(t + \Delta t)$  is the sum of the probabilities that (a) there are  $n$  customers at  $t$  and there is no arrival or completion of service in  $(t, t + \Delta t)$ , (b) there are  $n - 1$  customers at  $t$  and one customer arrives in  $(t, t + \Delta t)$ , and (c) there are  $n + 1$  customers at  $t$  and service is completed in  $(t, t + \Delta t)$ . Therefore

$$(2) \quad \frac{dp_n(t)}{dt} = -\left(\frac{1}{a} + \frac{1}{s}\right)p_n(t) + \frac{1}{a}p_{n-1}(t) + \frac{1}{s}p_{n+1}(t),$$

$$n = 1, 2, \dots$$

This set of simultaneous linear differential equations can be solved to give the transient behavior of the system. The answer is rather complicated. If it is assumed that an equilibrium distribution exists, the equations must have a solution,  $\{p_n\}$ , that is independent of time. Thus, directly from (1) and (2),

$$0 = -\frac{p_0}{a} + \frac{p_1}{s},$$

$$0 = -\left(\frac{1}{a} + \frac{1}{s}\right)p_n + \frac{1}{a}p_{n-1} + \frac{1}{s}p_{n+1},$$

$$n = 1, 2, \dots$$

Also, since the  $\{p_n\}$  form a probability distribution,  $1 = p_0 + p_1 + p_2 + \dots$ ; the solution is

$$\begin{aligned} p_0 &= 1 - r, \\ p_1 &= r(1 - r), \\ &\vdots \\ p_n &= r^n(1 - r), \\ &\vdots \end{aligned}$$

where  $r = s/a$  is the traffic intensity. The mean of the distribution is  $r/(1 - r)$ . When  $r = 0.9$ , the case examined previously, the mean number in the system is 9; this is the average over a long time of

the number of customers present. Because arrivals are random, an arriving customer also finds on the average 9 customers ahead, each taking on the average 0.9 min. to be served. Thus the mean time spent queueing, for the queue discipline first-come, first-served, is  $9 \times 0.9$  min. = 8.1 min. The general formula is that in a single-server queue, with random arrivals, exponential service time, and the queue discipline first-come, first-served,

$$(3) \quad \frac{\text{mean time spent queueing}}{\text{mean service time}} = \frac{1}{1 - r},$$

where  $r$  is the traffic intensity.

The method used above to get linear equations can be used for more complex systems, provided that arrivals are independent and random and that service time distributions are exponential. However, consider even the simple single-server queueing system with random arrivals but with a non-exponential distribution of service time; here the argument fails, because the probability that service of a customer is completed in  $(t, t + \Delta t)$  is not constant but depends on how long service has been in progress. There are various ways round this difficulty, of which, one—the method of the imbedded Markov process—depends on considering the system only at instants at which service of a customer is completed. The generalization of (3) is the Pollaczek-Khinchin formula,  $\frac{1}{2}r(1 + c^2)/(1 - r)$ , where  $c$  is the ratio of the standard deviation of service time to the mean.

*Future developments.* The number of new queueing situations is unlimited; almost each new application shows a fresh combination of input, service mechanism, and queue discipline. Simulation is usually the best method for tackling the more complex situations, and there is scope for further work on increasing the efficiency of simulation techniques. Two fields where further mathematical work is likely are those of networks of queues and of the automatic control of queueing systems.

*History.* The first major theoretical investigation was made in 1909 by A. K. Erlang, Copenhagen Telephone Co. Between the two world wars, important mathematical work on queueing was done by T. C. Fry (United States), A. Ia. Khinchin (Soviet Union), and Felix Pollaczek (France). After 1945, interest in the subject grew rapidly, both because of the much increased attention paid to operations research, and because the appropriate mathematical tool, the theory of stochastic processes, was being widely studied in other contexts. There is now an extensive literature on the mathe-

mathematical problems connected with queueing. Except for the special fields of telephone engineering and traffic studies, relatively little has been written about applications.

D. R. Cox

[See also MODELS, MATHEMATICAL.]

#### BIBLIOGRAPHY

*Introductory theoretical books in English are Khinchin 1955, Morse 1958, Cox & Smith 1961, and Riordan 1962. Saaty 1961 gives a more extensive review of the subject and has a long bibliography. Takács 1962 and Beneš 1963 are more specialized and mathematical. Syski 1960 deals particularly with applications to telephone engineering. Empirical and theoretical work on road traffic problems is best approached through the publications of the Road Research Laboratory (Great Britain) and the Highway Research Board (United States). White 1962 contains a critical discussion of work on the somewhat related problem of the sizes of casual groups of people.*

BENEŠ, VÁCLAV E. 1963 *General Stochastic Processes in the Theory of Queues*. Reading, Mass.: Addison-Wesley.

COX, DAVID R.; and SMITH, W. L. 1961 *Queues*. London: Methuen; New York: Wiley.

FLORIDA, UNIVERSITY OF, GAINESVILLE, STATISTICAL LABORATORY 1956 *Symposium on Monte Carlo Methods, Held at the University of Florida . . . March 16 and 17, 1954*. Edited by Herbert A. Meyer. New York: Wiley.

KHINCHIN, ALEXANDR IA. (1955) 1960 *Mathematical Methods in the Theory of Queueing*. London: Griffin; New York: Hafner. → First published in Russian.

MORSE, PHILIP M. 1958 *Queues, Inventories and Maintenance: The Analysis of Operational Systems With Variable Demand and Supply*. New York: Wiley.

RIORDAN, JOHN 1962 *Stochastic Service Systems*. New York: Wiley.

SAATY, THOMAS L. 1961 *Elements of Queueing Theory, With Applications*. New York: McGraw-Hill.

SYSKI, RYSZARD 1960 *Introduction to Congestion Theory in Telephone Systems*. Edinburgh: Oliver & Boyd.

TAKÁCS, LAJOS 1962 *Introduction to the Theory of Queues*. New York: Oxford Univ. Press.

TANNER, J. C. 1961 *Delays on a Two-lane Road*. *Journal of the Royal Statistical Society Series B* 23:38-63.

WHITE, H. 1962 *Chance Models of Systems of Casual Groups*. *Sociometry* 25:153-172.

#### QUOTA SAMPLING

See SAMPLE SURVEYS, article on NONPROBABILITY SAMPLING.





# R

## RACE

*The article under this heading discusses both the history of the concept of race and the cultural definition of race. The biology of race is described under GENETICS, especially in the article on RACE AND GENETICS. Social, psychological, and economic aspects of race are described in DISCRIMINATION, ECONOMIC; ETHNIC GROUPS; INDIVIDUAL DIFFERENCES; INTELLIGENCE AND INTELLIGENCE TESTING; MINORITIES; PREJUDICE; RACE RELATIONS; SEGREGATION. Also relevant is the article SOCIETAL ANALYSIS. The biography of Boas should be consulted for its discussion of Boas' pioneering studies of the diversity of cultural forms that occur within and between so-called racial groups.*

Race engages the attention of social scientists both as a special constellation of cognitive or ideological categories and as a means of explaining sociocultural phenomena. Race is a recurrent ingredient in the ethnosemantics of group identity and intergroup relations. "Social races" are composed of socially defined and significant groups; the study of social race thus is a fundamental aspect of the study of social structure, especially in stratified state societies. Raciological (biological) explanations of sociocultural differences and similarities have attracted the support of a large and not infrequently preponderant scientific consensus. Evaluation of the relevance of racial differences to sociocultural theory thus becomes an inescapable obligation of the social sciences.

### Social race

The term *race*, or its various ethnosemantic glosses, is applied in vernacular contexts to human

populations organized along an astonishing variety of principles. Nation-states, such as the Irish, Japanese, or German; tribes such as the Scythian, Iroquois, Zulu; language families such as Slavic, Latin, Semitic; minorities such as the Jews, gypsies, Puerto Ricans; and phenotypically distinct but genetically hybrid aggregates such as whites, Negroes, yellows, and Coloureds are cognitively equivalent in many ethnosemantic contexts. Social scientists have tried to diminish intergroup conflict by exposing the disparity between biologically acceptable definitions of race and those which are entertained at a popular level. Since none of the folk usages is informed by valid genetic principles, the lack of correspondence between social race and biological race should occasion no surprise. Social races encompass both phenotypically similar and phenotypically dissimilar populations; actual gene frequencies, the ultimate goal of infraspecies systematics, are obviously not desiderata in folk taxonomies. The cognitive substratum by which so many disparate aggregates are united cannot therefore be regarded as racial in the biological sense. It is quite obvious that in some societies biologically similar populations may be classified into groups which elsewhere go unrecognized. In this respect, the cognitive treatment of racial identity in Brazil is especially instructive. Despite the presence of distinct Negroid and Caucasoid components, the racial categories recognized by Brazilians are not the same as the categories recognized in the United States. In distinguishing socially defined races, therefore, attention must be directed toward common sociocultural rather than common biological features.

Viewed in terms of social structure, the processes by which a group achieves its continuity and by which individuals identify themselves and



others are especially significant. Social races are composed of subjectively significant groups, unrestricted by age and sex criteria, in which membership is sociocentric (i.e., appears the same to all egos), is established at birth, endures for life, and confers special behavioral obligations or privileges. Social races differ from other stratified groups (such as classes with low rates of out-mobility) in their methods of maintaining membership and group identity. Social races accomplish this by a special ideological device, the idea of descent. Although the members of a social race are replaced during each generation, the group maintains a continuing identity through varied applications of descent rules. The basic cognitive factor of social races thus is present wherever the decisive test of group membership involves the calculation of genealogical relationships.

Since all kinship groups identify their membership in part through recognized principles of descent, there is a degree of overlap between kin groups and social race. Unilinear descent groups, in which membership is restricted exclusively to consanguineals, exhibit most of the essential ideological components of folk "races"; because membership is determined by descent, it is unambiguous, endures throughout the life of each member, and is associated with behavioral specialties such as marriage restrictions, rituals, and jural privileges. Unilinear descent groups differ from the more familiar instances of social "races" such as the Jews, gypsies, and Negroes in displaying more harmonious relations with outgroups which are founded on exogamic exchanges. Such differences do not distinguish the case of Indian castes, which in all essentials, including intergroup relations, resemble the classic social races. The composite band and the kindred and other nonunilineal groups in which affiliation is regularly achieved through marriage, voluntary recruitment, or option show a correspondingly more remote resemblance to the classic social races.

Racially heterogeneous societies need not be divided into structurally significant social races. If a descent rule is not evolved, categorizations of hybrid phenotypes will tend to correspond to the actual variety of visible intergrades. This is the case in Brazil, where descent plays a negligible role in establishing "racial" identity. In Bahia, full siblings whose phenotypes markedly differ from each other are assigned by different members of their community to contrastive racial categories. Because of the absence of descent rules, pronounced disagreements concerning the identity of individuals frequently occur, and this ambiguity

is compounded by the plethora of "racial" categories—as many as forty in a single community and four or five hundred for the country as a whole. Under these conditions, structurally significant sociocentric groups based on racial criteria can scarcely be said to exist.

This conclusion is further warranted by the absence of systematic segregation and discrimination (but not of prejudice) associated with the social categories. Blurred, egocentric, and ambiguous racial identity would thus appear to be a functional correlate of the structural inconsequence of racial factors and the insignificance of descent (Harris 1964, pp. 54–64).

The importance of descent in racial ideology is frequently obscured by the apparent adequacy of phenotypical cues for the establishment of unambiguous identity. Most Americans are skilled in detecting phenotypical traces of Negroid admixture among the members of the community. Yet the much-discussed "high visibility" of the Negro group is belied by the phenomenon of the "white Negro," a self-identified Negro-by-descent who refuses to "pass" into the white group. The powerful fictions invented through application of a descent principle are further demonstrated by the treatment of children of mixed marriages. Regardless of phenotype, such children are affiliated with the racial group of the lower-ranking parent. This affiliation is not at all a "natural" reflection of phenotypical character; it is rather a social fiction made possible by the peculiar principles of the descent rule practiced in the United States. One of the most common methodological blunders in scientific studies of the significance of racial differences in the United States is the tacit acceptance of this phantasmic notion of race as the basis for establishing research samples (Fried 1964).

### Biological race

The development of scientific raciological theories is mainly a by-product of the development of Euro-American biology, although Carl von Linné's correlation of custom and race in his epoch-making taxonomy, *Systema naturae*, was certainly not without both learned and popular precedent. Whether the racism that can be discovered in the texts of antiquity, in the Babylonian Talmud, and in the writings of a few Greeks and many Romans deserves to be detached from the underlying matrix of folklore depends on how much emphasis one gives to the distinctive features of post-Galilean science. In Linné's classification of human varieties (1735), the aboriginal American is reddish in color with straight, dense black hair, and so on.

He is "persevering, content, free," and he paints himself with "skillful red lines." Europeans are "light, active, ingenious" and "covered with tailored clothes." Asians are "severe, haughty, miserly" and "covered with loose garments"; while Africans are "crafty, lazy, negligent, anointed with oil" and "governed by whim" (Count 1950, p. 359). Although Linné probably regarded these racial traits as having some degree of permanency, eighteenth-century biologists generally believed that only the species level of differentiation reflected true hereditary fixity. Linné, like others of that period who believed in the unitary origin of mankind, was obliged to leave the door open for evolution, because if all men were descended from Adam, infraspecies variation had somehow to be explained. Moreover, it was at this time that environmentalism, both social and natural, was being enshrined in radical bourgeois political programs under the influence of Locke's doctrine of the *tabula rasa*. Indeed, one of the central ideological themes of the French Revolution was the belief that hereditary differences among men were subordinate to the conditioning influences of education and the natural environment. Hence, on the whole, the eighteenth-century monogenesists tended to deprecate the power and permanence of racial differences. Buffon, for example, whose evolutionist inclinations appear at times to embrace the idea of the transformation of species, considered climate and diet to be the principal causes of racial variation. Like Linné, he correlated cultural and psychological characteristics with racial differences, yet so strong was his environmentalism that he believed all such differences would "gradually disappear . . . if the causes which produced them should cease, or if their operation should be varied by other circumstances and combinations" (Count 1950, p. 15). This early version of Lamarckism is conspicuous also in the outlook of men like Rousseau and Monboddo, both of whom were convinced that the great apes were actually human beings whose progress toward civilization had been retarded by adverse environmental conditions (Rousseau 1775, pp. 22 ff.; Monboddo 1773, pp. 22 ff.). Others, such as Blumenbach, one of the most influential naturalists of the period, continued to support the doctrine of "perfectibility" of the members of nonwhite races. In fact, belief in environmentalism and perfectibility was so strong that respectful attention was given to the possibility that Negroid physical features were not hereditary. In the United States, Samuel Stanhope Smith even argued that dark skin color was a phenomenon similar to freckles and that white people would

become Negroid with sufficient exposure to the sun. Smith illustrated the reverse process in the celebrated case of Henry Moss, a Virginia slave who appears to have suffered a loss of pigmentation after having moved to the North. At a meeting of the American Philosophical Society in 1797, Benjamin Rush argued that the Moss case provided proof that black skin color was a disease akin to leprosy (Gossett 1963, p. 41).

A substantial, perhaps dominant, body of scientific opinion in the eighteenth century was thus heavily committed to the belief that racial differences were rather evanescent and subject to the control of both natural and cultural aspects of the environment. It was this negation of racial determinism that permitted Turgot, Holbach, Helvétius, and Condorcet to attempt social analyses which foreshadowed the modern cultural-ecological approach to the understanding of sociocultural evolution. But the belief that race was relatively unimportant to an understanding of the variety of cultures, past and present, was not destined to flourish in the succeeding century.

The equalitarian philosophy of the monogenesists was based on faith in the power of the environment to modify Hottentots into Englishmen. Indeed, equalitarian monogenesists long before Lamarck were quite prepared to believe that a specific transformation from ape to Caucasoid man must have taken place as part of the process by which humanity had emerged from its "state of nature." A necessary but fatal ingredient in these early environmentalist theories was a stanch belief in the inheritance of acquired characteristics. Lamarck merely generalized this point of view to include the entire organic universe. His theory of the use and disuse of organs was in most fundamental aspects perfectly anticipated in the work of Rousseau and Monboddo (Lovejoy 1933). Moreover, there was no significant disagreement between Darwin and Lamarck on the power of the environment to shape the hereditary nature of bioforms. According to Darwin, the use and disuse of organs exerted a direct influence upon the innovations which are subjected to natural selection (Darwin 1859, pp. 133-137 in 1958 edition).

Yet Darwin, Wallace, Huxley, Haeckel, Spencer, and every other evolutionist of the late nineteenth century, as well as almost every major social scientist from Marx to Morgan, regarded racial differences as essential to the understanding of human behavior (cf. Zirkle 1959, pp. 109-111; Resek 1960, p. 63). Even before them, the polygenesists, influenced by the skepticism that led them to reject the theologians' idea of "special creation,"



tended to regard the races as permanent, species-like divisions, with differential hereditary capacities for achieving civilization. Voltaire, the influential Lord Kames, the English physician Charles White, the Frenchman Virrey, and the American Jefferson were among those who emphasized inequality of the hereditary endowments of the human races and the importance of these differences for an understanding of sociocultural phenomena (cf. Gossett 1963, pp. 42-53). Although post-Darwinian biologists tended to reaffirm the monogenesist view of a unitary origin for all mankind, they drew none of the equalitarian conclusions that had characterized their eighteenth-century predecessors.

What had happened to turn a doctrine of equality, optimism, and perfectibility into chauvinistic, imperialistic, competitive, and pessimistic racism? One major scientific development of the early part of the nineteenth century seems especially significant (not to deny that a more fundamental answer would involve social, as distinct from ideological, innovation). A dawning insight into the true dimension of geological time compromised the equalitarianism of the evolutionists and brought them into direct conflict with the Biblically inspired equalitarian monogenesists. Hitherto unimaginably prolonged sequences were revealed, giving a completely new perspective to the transformation of species. Savages could be turned into Englishmen, yes, but not overnight, or in a few generations. According to the new chronology of progress, Caucasoid man came to be regarded as thousands of years "ahead" of the other races. The latter remained perfectible, but along a time scale which effectively removed them from any immediate claim to equal sociopolitical treatment. Characteristics were still acquired and became part of the hereditary repertory, but at a painfully slow pace. The ideological inertia engendered by concepts of racial heritage, racial memory, and racial consciousness enveloped the social and biological sciences and cast a spell over vast domains of Euro-American arts and letters. In its most popular version, promulgated principally by the synthesizing genius of Herbert Spencer, the social-to-biological feedback operated through the mechanisms of natural selection and the survival of the fittest: fundamental sociocultural change required equally fundamental biological change. Survival in an environment of unrestricted competition required individualized struggle and the untrammelled exercise of the will to achieve power. Hence, political philosophies such as social Darwinism, the "white man's burden," and "manifest destiny" all moved toward the mass megalomania of German Ary-

anism and twentieth-century race war and genocide (Barzun 1937; Hofstadter 1944; Snyder 1939; Gossett 1963). Not to be omitted from this spectacle is the contribution of the caliper-wielding anthropometrists, from Topinard to Lombroso, with their obsession for quantifying the precise details of physiologically insignificant traits. With slight consideration of the need for experimental control over cultural and physiological variables, minute differences, such as in the angles formed by cranial and facial bones, were used to help explain sociocultural contrast (Boyd 1958). Against the combined weight of the bioevolutionary synthesis, the passionate identification of science with precise measurement, and the problem of reconciling the inequalities of imperialism and capitalism with Christian principles, few intellectuals were able to withstand the appeal of some form of scientific racism.

Adherence to this ill-fated conceit was further promoted by an overemphasis, among the great cultural evolutionists of the times, upon parallel evolution and convergence. The simplistic equation of graded physical types with graded cultural types was extremely vulnerable to the test of comparative observation.

Documentation of the vast diversity of cultural forms that occur within as well as between racial groups eventually severed racist explanations from every vestige of scientific respectability. Much of the credit for this achievement rests with Franz Boas and the historical particularists and cultural relativists who were nurtured under his tutelage. Boas and his students insisted that there was no causal relationship between race, language, and culture (Boas 1887-1936; 1911). Especially damaging to the racist point of view were the first sophisticated linguistic descriptions of present-day primitive peoples, whereby it was established that variability and complexity of grammar and syntax are totally unrelated to degree of technological or sociopolitical organization (Sapir 1921).

The distinctive feature of the modern consensus concerning the relationship between race and culture is that the rate and direction of culture change among the various infraspecies groupings of *Homo sapiens* is not at present significantly affected by genetic specialties. It is generally believed that if all factors other than race are held constant, similar enculturation experiences will result in similar sociocultural repertories. Indeed, if perfect control over the enculturative process could be achieved, one generation would suffice to equip any two *Homo sapiens* groups with essentially similar repertories, regardless of their respective racial phy-

logenes. Thus, if Hottentot babies were substituted at birth for English babies, their average cultural performance would probably not differ in any significant fashion from a control group. The evidence for this point of view is overwhelming. Social groups and individuals drawn from every human "race" have shown themselves in countless instances to be susceptible to acculturation influences bearing upon every aspect of sociocultural behavior. American Indians brought up in Brazil show no hereditary resistance to learning African dance rhythms; Englishmen reared in China can learn to speak flawless Chinese; American Negroes who attend conservatories write symphonies in the classical European tradition; the Japanese display not the slightest hereditary disability in acquiring a knowledge of Western electronics; Jews brought up in Germany have German food preferences, while those brought up in Yemen acquire Yemenite tastes; under the influence of Western missionaries, the South Sea islanders learned to govern their sexual affairs in conformity with strict Protestant codes; everywhere the children of illiterate parents, exposed to the proper set of enculturative conditions, acquire within one lifetime the learning and lore contributed by hundreds of generations of men from all the races of the world. Although it is not possible to prove that all large divisions of *Homo sapiens* have equal learning ability for all kinds of responses, it is beyond dispute that the overwhelming bulk of the response repertory of any human breeding population can be learned by any other human breeding population. Moreover, if average differences in learning ability exist, they are demonstrably insufficient to account for the major cultural and subcultural contrasts which occupy the attention of the social sciences.

No one familiar with modern ethnography can doubt the dominant role of enculturative conditioning in the establishment of the behavioral specialties manifested by different human populations. No plausible connection has ever been proposed between specific human genes and such specialties as cross-cousin marriage, bilateral descent, polyandry, divine kingship, monotheism, brideprice, private property in land, or thousands of other small and large chains of less than universal human behavior. On the other hand, the inadequacy of racist explanations of sociocultural differences and similarities is further underscored by the growing success achieved by cultural and ecological explanations of those phenomena. With the exception of a handful of hereditary pathological disabilities, there does not exist a single instance of differential learning ability between populations which cannot

be explained in terms of differential conditioning experience. Certainly this is the case wherever scores on so-called intelligence tests have been correlated with "racial" groups. Over and over again it has been demonstrated that such scores reflect, for each individual, the number of years of schooling, the quality of his academic training, his specific training for the test situation, his nuclear and extended family milieu, and an abundance of other nongenetic conditioning parameters (Klineberg 1935; 1951; 1963; Comas 1961; Brown 1960; Dreger & Miller 1960). As long as average performance differences are demonstrably embedded in conditioning parameters for which plausible linking hypotheses exist, emphasis upon genetic factors, especially those not significantly related to the behavior traits in question, exposes social science to every sort of occult mystification through the multiplication of superfluous hypotheses.

It should be carefully noted that the consensus by which racist explanations in the social sciences have been deprived of all claim to scientific respectability does not diminish the importance of further research concerning the essentially "creative" relation between biological and cultural traits. Such considerations are obviously required for an understanding of the evolution of the human condition, both during the hominoid-hominid transition and during the periods of species changes among the early hominids. Moreover, it is well known that human breeding populations exhibit genetic specialties, such as achondroplasia, myopia, and diabetes, whose frequencies are affected by cultural factors. Certainly, the rejection of racism should not obscure the possibility that differential reproductive and survival rates account for some cultural differences. None of these relationships, however, need run counter to the assumption that learning ability in its widest sense, as enculturation potential, is equal among all *Homo sapiens*. On the contrary, we may continue to assume that selection for behavioral plasticity, for freedom from gene-specific behavior, for high-order generalized learning receptivity, has been the most important source of biological control over the direction of hominid evolution during the last million and a half years (Dobzhansky & Montagu 1947; Hockett & Ascher 1964).

MARVIN HARRIS

#### BIBLIOGRAPHY

- BARNES, J. A. 1960 *Anthropology in Britain Before and After Darwin*. *Mankind* 5:369-385.  
 BARZUN, JACQUES (1937) 1965 *Race: A Study in Superstition*. Rev. ed. New York: Harper. → First published as *Race: A Study in Modern Superstition*.



- BENEDICT, RUTH (1940) 1959 *Race: Science and Politics*. New York: Viking.
- BERRY, BREWTON (1951) 1965 *Race and Ethnic Relations*. 3d ed. Boston: Houghton Mifflin. → First published as *Race Relations: The Interaction of Ethnic and Racial Groups*.
- BLUMENBACH, JOHANN F. (1775-1811) 1865 *The Anthropological Treatises of Johann Friedrich Blumenbach*. Translated and edited by Thomas Bendyshe. London: Longmans. → First published in German.
- BOAS, FRANZ (1887-1936) 1955 *Race, Language and Culture*. New York: Macmillan.
- BOAS, FRANZ (1911) 1963 *The Mind of Primitive Man*. Rev. ed. New York: Collier. → A paperback edition was published in 1965 by the Free Press.
- BOYD, WILLIAM C. 1958 Has Statistics Retarded the Progress of Physical Anthropology? *American Journal of Physical Anthropology* New Series 16:481-484.
- BROWN, INA C. 1960 [A Book Review of] *The Testing of Negro Intelligence*, by Audrey M. Shuey. *American Anthropologist* New Series 62:544 only.
- COMAS, JUAN 1961 "Scientific" Racism Again? *Current Anthropology* 2:303-340.
- CONFERENCE ON RACE RELATIONS IN WORLD PERSPECTIVE, HONOLULU, 1954 1956 *Race Relations in World Perspective: Papers*. Edited by Andrew W. Lind. Honolulu: Univ. of Hawaii Press.
- COON, CARLETON S. 1962 *The Origin of Races*. New York: Knopf.
- COUNT, EARL W. (editor) 1950 *This Is Race: An Anthology Selected From the International Literature on the Races of Man*. New York: Schuman.
- COX, OLIVER C. (1948) 1959 *Caste, Class and Race: A Study in Social Dynamics*. New York: Monthly Review Press.
- DARWIN, CHARLES (1859) 1964 *On the Origin of Species*. Cambridge, Mass.: Harvard Univ. Press.
- DOBZHANSKY, THEODOSIUS 1962 *Mankind Evolving: The Evolution of the Human Species*. New Haven: Yale Univ. Press.
- DOBZHANSKY, THEODOSIUS; and MONTAGU, ASHLEY 1947 Natural Selection and the Mental Capacities of Mankind. *Science* New Series 105:587-590.
- DOBZHANSKY, THEODOSIUS; and MONTAGU, ASHLEY 1963 Two Views of Coon's Origin of Races: With Comments by Coon and Replies. *Current Anthropology* 4:360-367.
- DREGER, RALPH M.; and MILLER, KENT S. 1960 Comparative Psychological Studies of Negroes and Whites in the United States. *Psychological Bulletin* 57:361-402.
- FRAZIER, E. FRANKLIN (1957) 1965 *Race and Culture Contacts in the Modern World*. Boston: Beacon.
- FRIED, MORTON H. 1964 Disastrous Definition: Comment on Dwight J. Ingle's *Racial Differences and the Future*. *Science* New Series 146:1526 only.
- GARN, STANLEY M. 1957 Race and Evolution. *American Anthropologist* New Series 59:218-224.
- GOSSETT, THOMAS F. 1963 *Race: The History of an Idea in America*. Dallas, Texas: Southern Methodist Univ. Press.
- HARRIS, MARVIN 1964 *Patterns of Race in the Americas*. New York: Walker.
- HOCKETT, CHARLES F.; and ASCHER, ROBERT 1964 The Human Revolution. *Current Anthropology* 5:135-168.
- HOFSTADTER, RICHARD (1944) 1959 *Social Darwinism in American Thought*. Rev. ed. New York: Braziller.
- HUGHES, EVERETT C.; and HUGHES, HELEN M. 1952 *Where Peoples Meet: Racial and Ethnic Frontiers*. Glencoe, Ill.: Free Press.
- KLINEBERG, OTTO 1935 *Race Differences*. New York: Harper.
- KLINEBERG, OTTO (1951) 1953 *Race and Psychology*. Paris: UNESCO.
- KLINEBERG, OTTO 1963 Negro-White Differences in Intelligence Test Performance: A New Look at an Old Problem. *American Psychologist* 18:198-203.
- KLOOSTERBOER, WILLEMINA 1960 *Involuntary Labour Since the Abolition of Slavery: A Survey of Compulsory Labour Throughout the World*. Leiden (Netherlands): Brill.
- KROEBER, ALFRED L. 1915 Eighteen Professions. *American Anthropologist* New Series 17:283-288.
- LINNÉ, CARL VON (1735) 1956 *Systema naturae*. 10th ed. London: British Museum.
- LIVINGSTONE, FRANK B. 1962 On the Non-existence of Human Races. *Current Anthropology* 3:279-281.
- LOVEJOY, ARTHUR O. 1933 Monbodo and Rousseau. *Modern Philology* 30:275-296.
- MACK, RAYMOND W. 1963 *Race, Class and Power*. New York: American Book.
- MONBODDO, JAMES B. (1773) 1774 *Of the Origin and Progress of Language*. 2d ed. Edinburgh: Balfour.
- MONTAGU, ASHLEY (1945) 1964 *Man's Most Dangerous Myth: The Fallacy of Race*. 4th ed., rev. & enl. Cleveland: World.
- PUTNAM, CARLETON 1961 *Race and Reason: A Yankee View*. Washington: Public Affairs Press.
- RESEK, CARL 1960 *Lewis Henry Morgan. American Scholar*. Univ. of Chicago Press.
- RITTERBUSH, PHILIP C. 1964 *Overtures to Biology: The Speculations of Eighteenth-century Naturalists*. New Haven: Yale Univ. Press.
- ROUSSEAU, JEAN JACQUES (1775) 1941 *Discours sur l'origine et les fondements de l'inégalité parmi les hommes*. Cambridge Univ. Press.
- SAPIR, EDWARD A. 1921 *Language: An Introduction to the Study of Speech*. New York: Harcourt.
- SIMPSON, GEORGE E.; and YINGER, J. MILTON (1953) 1965 *Racial and Cultural Minorities: An Analysis of Prejudice and Discrimination*. 3d ed. New York: Harper.
- SNYDER, LOUIS L. 1939 *Race: A History of Modern Ethnic Theories*. New York: Longmans.
- THOMPSON, EDGAR T. (editor) 1939 *Race Relations and the Race Problem: A Definition and an Analysis*. Durham, N.C.: Duke Univ. Press.
- THOMPSON, EDGAR T.; and HUGHES, EVERETT C. (editors) 1958 *Race: Individual and Collective Behavior*. Glencoe, Ill.: Free Press.
- UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION 1961 *The Race Question in Modern Science: Race and Science*. New York: Columbia Univ. Press. → Contains all the brochures that have appeared separately in the UNESCO series "The Race Question in Modern Science."
- WAGLEY, CHARLES 1958 On the Concept of Social Race in the Americas. Volume 1, pages 403-417 in *International Congress of Americanists, Thirty-third, San José, Costa Rica, 1958, Actas*. San José: Lehmann.
- WAGLEY, CHARLES; and HARRIS, MARVIN 1958 *Minorities in the New World*. New York: Columbia Univ. Press.

- WASHBURN, S. L. 1951 *The New Physical Anthropology*. New York Academy of Sciences, *Transactions* Second Series 13:298-304.
- WASHBURN, S. L. 1963 *The Study of Race*. *American Anthropologist* New Series 65:521-531.
- WILLIAMS, ERIC (1944) 1961 *Capitalism and Slavery*. New York: Russell & Russell.
- ZIRKLE, CONWAY 1959 *Evolution, Marxian Biology, and the Social Scene*. Philadelphia: Univ. of Pennsylvania Press.

## RACE RELATIONS

- I. WORLD PERSPECTIVES G. Franklin Edwards  
II. SOCIAL-PSYCHOLOGICAL ASPECTS Thomas F. Pettigrew

### I WORLD PERSPECTIVES

The term "race relations" refers to those forms of behavior which arise from the contacts and resulting interaction of people with varied physical and cultural characteristics. As defined by Robert E. Park (1939), the concept refers to all relationships which are capable of producing race conflict and race consciousness and which determine the relative status of groups in the community.

It should be noted that differences in physical and genetic traits are important in contributing to the observed ecological, economic, social, and political relationships which constitute the subject matter of race relations. However, since contacts among people of diverse racial origins have usually involved groups with markedly variant technologies, patterns of social organization, political systems, and religious beliefs and values, the biological aspect of race has usually been interpreted as reinforcing these other differences rather than as a primary, independent factor in the observed behavior. Expressed in other terms, the association of people belonging to different racial groups also involves the association of groups with different cultural characteristics.

**Related concepts.** A distinction should be made between the usage of the term "race relations," which refers to the social processes and social structures arising from the contacts of different racial groups, and other related usages. The term is variously employed to cover forms of intergroup, interethnic, and majority-minority relationships. In these latter usages, race may or may not be a significant variable in the behavior under question. In general, these cognate concepts, though often used synonymously with that of race relations, encompass other forms of behavior as well.

In the United States since World War II, the

concept "intergroup relations" has been employed widely to cover forms of organized group relations involving tensions, conflict, and group adjustments. These include, in addition to Negro-white relationships, the interaction of religious groups and of organizations representing different economic strata. The major analytical focus in this area of study is upon such social-psychological phenomena as prejudice, discrimination, power relations, leadership roles and strategies, and the impact these make upon group relations and community change.

Although the concept "ethnic relations" is often employed to cover what traditionally is studied as race relations, it is not limited to analyses involving people with different racial characteristics. Ethnic groups are separated from others among whom they live by distinctive characteristics which provide a consciousness of difference [see ETHNIC GROUPS]. Besides race, the differentiating features may be religion, language, nationality, or some combination of these, which provide the ethnic group with a strong in-group feeling. In parts of Canada, for example, conflict and acutely self-conscious behavior characterize the relationship between the English-speaking and French-speaking communities, whose people are of the same racial origin but vary in language, religion, national origin, and traditional behavior.

Minorities are ethnic or racial groups that occupy subordinate positions in the communities where they reside [see MINORITIES]. In addition to segregation from other members of the community because of some racial, social, or cultural characteristic, they often suffer severe political restrictions. Their status is characterized by accommodation. The rearrangement of political boundaries so that people with different cultural traits are brought within a common national state is one source of the origin of minority groups. The realignment of national boundaries in the Balkan states following World War I is a classic instance of this process. In more recent times, following the end of British colonial rule in India, the emergence, owing to religious differences, of India and Pakistan as separate states—the former predominantly Hindu and the latter Muslim—has witnessed the continued presence of Muslim minorities in India and Hindu minorities in Pakistan.

While these related terms can refer to behavior which is shaped by a wide variety of group characteristics, they are sometimes used to describe and analyze phenomena in which race does play an important part. In the United States, for ex-



ample, the relations of Negroes and whites are conceptualized variously as intergroup, interethnic, and majority-minority relations.

### European colonization

If we leave aside the early migrations of primitive peoples, the era of race relations can be said to have begun with the overseas expansion of the major European powers from the fifteenth century onward. Over the following centuries England, France, Spain, and Portugal colonized the Americas and, with the addition of Dutch and Scottish traders, the Far East. The early settlement of the Dutch in South Africa was followed in the nineteenth century by the division of the African continent among the major European powers.

Economic expansion was the primary, but not exclusive, motive for the establishment of foreign settlements and for other contacts of Europeans with indigenous peoples in various parts of the world. Following the industrial revolution it was the need for raw materials and markets which led to the exploitation of non-European labor in different parts of the world and to the forms of social and political control which characterized the relations of European and non-European peoples.

E. Franklin Frazier (1957) distinguished three major types of racial and cultural frontiers created by this overseas expansion of European civilization: (1) those in which Europeans founded rather substantial permanent settlements (the United States, Latin America—including Mexico, Central America, South America, the West Indies and the Guianas—South Africa, Australia, and New Zealand); (2) the tropical dependencies, where Europeans were unable to form permanent settlements on a substantial scale (Africa south of the Sahara, southeast Asia, and the Pacific islands of Polynesia, Micronesia, and Melanesia); (3) the older civilizations of Asia (China and Japan), which constitute a relatively recent frontier. The relationships of Europeans with non-European peoples varied according to the type of frontier. Philip Mason (1954) enumerated three factors as having an important influence upon the race relations pattern which developed in each area: the type of climate and territory, the level of the dominated race's civilization, and the attitude of the colonial power.

In the United States, where climate was propitious for white settlement, two very distinctive patterns developed as a result of variations in a complex of factors—land and climate, type of production, and labor requirements. In the northern section of the country, climate, soil type, and rain-

fall led to the development of small farm units, fishing, and small industries, the labor requirements of which could be met by family members or free workers. In the South, however, the climate and terrain favored the production of staple crops and a system of plantation agriculture, for which a more abundant supply of cheap labor was necessary. After unsuccessful attempts to enslave the native Indians and to use indentured workers, the labor problem was solved by importing several million Africans as slaves over a period of two centuries. The status of Negroes as slaves throughout this period and their continued residence in the South following emancipation help to explain why race relations in the United States have traditionally been regarded as a "southern problem."

A pattern similar to that noted in the American South may be observed in other parts of the world where plantation agriculture developed—Jamaica, Brazil, and Java, to name only a few areas, as well as in South Africa, where mining also became an important industry. These areas lie in the tropics or subtropics, where climate is not conducive to the heavy settlement of whites and therefore precludes the employment of free white labor on an extensive scale. The need to assure a stable labor supply resulted in the exploitation of native people through low wages, forced labor, or enslavement. In Latin America and the West Indies, as in the United States, slave labor was brought from Africa. After the abolition of slavery in the Western world in the nineteenth century, cheap labor was brought from India and China, and the exploitation of native workers was continued. Negroes in the United States continued to be exploited as farm tenants by their white landlords.

**Variation in colonial policies.** The variation in approach to race relations by the European powers may be represented here by the dissimilar attitudes and practices of the English and Portuguese toward native peoples in the Americas, and by the contrasting patterns of the French and Portuguese, on the one hand, and the English and Dutch, on the other, in Africa.

In the English colonies in the United States, the establishment of permanent settlements favored the importation of white women in substantial numbers. The colonists did not intermarry with the Indians and were socially distant from both Indians and Negroes. Although moral conflict was engendered by the effort to reconcile the tenets of Protestantism with slave holding, as slavery became institutionalized in response to increasing labor demands, it was rationalized and justified

by ideologies based upon the conception that the blacks were inferior. The argument most frequently made was that since black men were biologically inferior, slavery was a natural state that was of value to slaves and masters alike; the conception that the institution was ordained by God was often added. Following the emancipation of the slaves, social distance between the races was maintained by a variety of devices. Thus, in the South there developed a caste-like system which was supported by legislation and state constitutions; Negroes not only were regarded as socially unequal but also became the victims of economic and political discrimination.

The pattern of relations which developed in Brazil is in contrast with that of the United States. The Portuguese, although holding Negroes as slaves until 1888, manifested a different attitude toward both Negroes and the native Indians (Freyre 1933; Pierson 1942). As Roman Catholics, they proselytized to convert both slaves and Indians to their religion, intermarried with them, and, following the emancipation, recognized the potentiality of members of these other racial groups to assume important positions as Brazilians. Brazilian society today is characterized by a marked absence of prejudice and discrimination based upon color and thus has no race problem of the type found in the United States.

The colonial policy of France and Portugal with regard to their African territories was one of eventual assimilation. Whenever educational and economic achievement justified it, Africans in the French and Portuguese colonies could expect to transfer from the status of subjects to that of citizens and to be governed by the laws of the metropolitan community rather than by the customary sanctions of local councils. In general, there was never any serious consideration of independent status for the colonies of these countries; it was expected that in time they would become a part of Greater France or Greater Portugal.

The English approach did not contemplate the evolution of African subjects, or those of the southeast Asia territories, to the status of citizens of the mother country. In contrast with the policies of France and Portugal, the long-range English objective was the independence of colonies as they became ready for self-government. In most English colonies the practice of indirect rule was followed. Parallel institutions were developed for the governance of Englishmen and native subjects, and native Africans were governed by their own institutions, with native rulers accountable to repre-

sentatives of the English government. Under such an arrangement, there was little social mixing, and discrimination in wage payments to native laborers and free white laborers was practiced.

The extreme expression of separation of colonists and natives occurred in South Africa, where, during the early history of the colony, the Dutch Afrikaners, with the encouragement and sanction of the Dutch Reformed church, developed a policy of rigid segregation of the native Hottentots and Bantus. Later, the East Indians who came to the area and the Cape Coloured people—a hybrid population resulting from the mixture of Afrikaners and natives—were also subjected to social and moral isolation. The segregation of colonists and natives is expressed in residential patterns; under the present-day policy of apartheid, native peoples are forced to live in designated compounds, apart from the descendants of the European settlers. Even in the large urban areas such as Durban (Kuper et al. 1958), the areas where non-Europeans may live are delimited. The English in South Africa have accommodated their policies to accord with those of the politically dominant Dutch party, especially with regard to segregation in social life. The English approach, however, is tempered by the humanitarian approach of missionaries and the more cosmopolitan views of the English in general, so that it is less restrictive than that of the Afrikaners in the provision of education and welfare services.

The approaches described as characteristic of the French and Portuguese, and those of the English and Dutch, must be taken as ideal descriptions. In reality, some variation in practice existed among each of these powers from one colony to another. The French in Algeria, for example, prior to the independence of that country, departed to a considerable extent from the assimilationist doctrine, owing mainly to the presence of large numbers of Muslims in the Algerian population. The English in Kenya, it should be observed, were much less restrictive toward the natives than they were in South Africa.

### Patterns of contact

In those parts of the world where Europeans settled or had extensive contacts with native peoples, a number of fairly uniform patterns and changes may be observed. One of these is demographic. The presence of Europeans exposed natives to a large number of infectious diseases to which they had little or no immunity, such as tuberculosis, syphilis, smallpox, and measles.



Heavy losses of population from these diseases occurred among the native Indians of the Americas and the West Indies, the Maoris of New Zealand, and native populations in Australia and other countries.

After the native population acquired some immunity and public health measures and medical care had been improved, population decline was arrested by the lowering of mortality rates. There followed a period of sharp population increase as birth rates remained constant or increased and mortality continued to be reduced. The Indian population of the United States (now a little over half a million, according to the 1960 census) has, after suffering heavy losses from disease and warfare, increased to a point approximating the estimated number during the pre-Columbian period (Carr-Saunders 1936, p. 300; compare Palmer 1948, p. 261). Among the Maori and other natives of New Zealand and Australia the same tendencies are observed (Price 1950, chapters 8 and 9; compare Carr-Saunders 1936, p. 296). Where native populations have increased, there is often a struggle on their part, especially in plantation areas, for an adequate food supply. Pre-emption of much of the best land by Europeans has resulted in food shortages and an acute struggle for existence.

Although disease was a primary contributor to the decimation of population among native peoples during the early period of their contacts with whites, it was not the only cause of population decline; the introduction of firearms also played a part. In some instances, as in encounters with Indian tribes of the Americas, the Tasmanians, the Hottentots of South Africa, and numerous other groups, the European settlers used firearms to inflict severe losses upon the natives before conquering them. Firearms, moreover, made warfare among native groups more destructive and contributed to the more extensive killing of wild animals, an important source of food. The suffering experienced by the Plains Indians of the United States as a result of the virtual elimination of the buffalo illustrates the point.

A second observed phenomenon growing out of the contacts of Europeans and native peoples is the formation of mixed-blood, or hybrid, populations. Wherever the races have met, miscegenation has occurred and new population types and culture groups have been formed. A classic instance of race mixing has occurred in Hawaii, where, over time, the native Hawaiians of Polynesian extraction, Portuguese, Chinese, Japanese, Puerto Ricans, Filipinos, Koreans, Negroes, and other Caucasian and Oriental people have inter-

married or otherwise mixed to produce new population types (Adams 1937; Lind 1938).

Race mixture has produced people who are marginal in both their physical and cultural characteristics—hence the term “half-caste” (Dover 1939). On the southeast Asian and Pacific frontiers the mixture has resulted in various Eurasian groups; in South Africa the Cape Coloured are the product of the admixture of the Dutch Boers and native Hottentots; in Latin America the union of the Spanish and the Portuguese with the Indians of the South American mainland resulted in the mestizos; and in the United States the mulatto population is the product of the mixture of white Americans and Negro slaves. These marginal populations not only are physically different from the parent populations but they also live in two cultural worlds. They are not fully accepted by the Europeans, and they, in turn, do not fully identify with the natives. Their intermediate position in the social structure makes them extremely self-conscious (Stonequist 1937).

The argument that these marginal people are biologically superior to natives (Reuter 1918) has been rejected, not only because it runs counter to genetic findings but also because it can be shown that the alleged superiority of these groups over indigenous populations results from greater access to available opportunities and social contacts (Wirth & Goldhamer 1944, p. 335). In the United States, for example, many of the mulattoes became free persons, chiefly through manumission, prior to the general emancipation of the slaves in 1863; thus they were able to become literate, purchase property, and develop stable families. With the formation of Negro communities following the emancipation, the mulatto was able to occupy leadership roles in politics and the professions. Since the large-scale urbanization of the Negro population and the resulting broadened opportunities for education, the mulatto no longer enjoys a virtual monopoly over status positions in the Negro community (Edwards 1959).

A third outgrowth of the contacts of Europeans and native peoples is the modification of the social organization of the latter. In general, the Europeans' superiority in technology and scientific knowledge resulted in new methods of cultivation and work relationships which served to change the prevailing pattern of most native groups. In those areas where industrialization has developed and a wage system of payment is used, the communal relationships and close kinship ties of native people built around agriculture have been undermined (Wilson 1936).

European values and ideologies have had an impact on the social relationships of native people and have produced considerable conflict and alienation. In the diffusion of such values, the missionaries, who lived in intimate relationship with the natives, played a significant role. Their teaching of Christianity to native groups introduced an ethic of individualism which disoriented converts from customary communal ties and often produced intratribal conflicts by challenging the moral basis upon which tribal life rested (Schapera 1940). The provision of welfare services often served a similar end. The mission schools made many students disenchanted with native life, and the introduction of modern scientific health practices and beliefs seriously challenged native magic and witchcraft.

Some students have abstracted from the history of race relations a sequence of changes which occurs from the time of initial contact until assimilation or some form of accommodation is reached. Park (1926), for example, in his "race relations cycle" theory, refers to the following pattern which he thought progressive and irreversible: initial contact, competition, accommodation, and eventual assimilation or amalgamation. It should be noted, however, that although most students recognize that uniformities and similarities do exist, there is nonetheless a disposition to question attempts to develop a patterned sequence which applies to all situations in which different racial and cultural groups have associated (Berry 1951).

### Contemporary problems

Contemporary interest in race relations centers on developments since the end of World War II, particularly on events in Africa and in the United States. The concern with race relations today, however, is universalistic and transcends local areas and the particularism of the people involved. The United Nations, founded in 1946 as an instrument of international cooperation and peace, is dedicated to the principles of equal rights and self-determination of peoples. These principles nurtured the rising tide of anticolonialism and nationalism which, during the following two decades, resulted in the emergence of many new nations in Africa and the Near East; 32 of these emergent nations are in Africa south of the Sahara.

The new relationships between the white people of the former colonial powers and the colored people of the new nations are, of course, matters of race relations as well as of world economics, politics, and geography.

The transformation from former colonial ter-

ritories to nation-states has produced changes in the social structures of the new states, although not everywhere in the same fashion. There has developed, moreover, a sense of power and national identity as black men serve as heads of state and succeed to important positions formerly occupied by representatives of the colonial powers. This sense of power and identity is expressed in a highly self-conscious literature which began as an expression of protest against colonial rule. A foremost aspect of this literature is an emphasis upon the concept "Negritude," which was developed initially and most fully by the writers of French West Africa. "The theory of Negritude," Balandier (1956, p. 150) points out, "appears as an exaltation of African-Negro specificity, as a kind of highly elaborated counter-racism."

South Africa. While most of the countries of west and east Africa have become independent, events have moved in an opposite direction for native peoples of South Africa. In the Republic of South Africa, the policy of apartheid was established in 1948 by the Nationalist party, which is devoted to the doctrine of white supremacy and which has been in power continuously. Apartheid, in essence, has the objective of maintaining the domination of three million whites over eleven million natives through physical separation of the races, both in residential areas and in places of public accommodation.

In 1959, the apartheid policy was extended by the passage of the Bantu Self-Governing Act. On the theory that each group could develop to its fullest capacity in its own homeland and that the Bantus were not a homogeneous people, eight reserves, or Bantustans, were established as homelands for the natives. Three million native Africans who lived in urban communities, some of whom were three generations removed from native life, were no longer regarded as permanent residents of such areas. They were, rather, to be considered as belonging to Bantu areas and would be interchangeable with residents of the Bantustans as the labor market required.

Although the restrictive character of life in South Africa is rationalized in terms of "separate development" and "peaceful coexistence" of the races, its results have severely penalized the non-white groups through land limitations, restrictions on freedom of movement, and differential voting privileges and expenditures for education, health, and other welfare measures. Although apartheid has been criticized within the republic by the National Union of South African Students, the South African Institute of Race Relations, the Women's



Defense of the Constitutional League (the Black Sash), and the Christian church (excluding the Dutch Reformed church), it has been rigorously enforced. There have been some riots by natives in protest against this regime, but so far such efforts have resulted only in heavy losses of life. To many observers, South Africa is a police state (Bunting 1964).

Racial conditions in South Africa are so much at variance with the movement toward freedom and independence of subject peoples throughout the world that they have become the object of world discussion and reaction. Discussion of the problems has been carried on in the United Nations almost continuously since 1946, when the Indian government protested the treatment of Indians in the Union under the Asiatic Land Tenure Act and the reprisals which followed their passive resistance campaigns. The question of South Africa's mandate over South-West Africa also came before the United Nations in 1946, and it has been a matter of continuous concern since 1948, when apartheid policies were introduced into the territory by the Nationalist government. In 1960 Liberia and Ethiopia petitioned the International Court of Justice, asking the court to find that South Africa had violated its obligation in administration of the South-West African territory and requesting that an order be entered to have discriminatory practices cease. In 1966 the court ruled that it had no jurisdiction in the matter. This decision was unexpected, in view of the careful legal preparation that had gone into the petition, which was supported by the United States.

**United States.** In the United States much progress has been made since the end of World War II in improving the position of racial minorities. Although the spotlight has been on the status of Negro-Americans, who constitute the largest and the most active group working for change, Mexican-Americans, American Indians, and Orientals have also benefited from antidiscriminatory legislation.

In the 1880s and 1890s, segregation became legalized in the South through the passage of various "Jim Crow" laws and changes in state constitutions which were subsequently upheld by Supreme Court decisions; for example, in the precedent-setting case of *Plessy v. Ferguson* (1896) the Court held that state laws requiring racial segregation were constitutional so long as "separate but equal" facilities were provided. Separate institutions were operated for the two major racial groups, separate facilities were maintained in

places of public accommodation, racial intermarriage was forbidden, and the right of suffrage was effectively nullified through legislation and practice. Thus the rights guaranteed Negroes through passage of the thirteenth, fourteenth, and fifteenth amendments to the constitution in the period between 1865 and 1870 were undermined. Furthermore, in the northern cities, to which Negroes migrated in large numbers during World War I and immediately thereafter, large Negro ghettos were formed as a result of discriminatory practices in housing and employment. While segregation in the southern states was based upon both law and custom, in the North *de facto* segregation resulted from discriminatory actions of groups and individuals.

In the 1930s the National Association for the Advancement of Colored People initiated a series of court cases asking that facilities for Negroes be made equal to those provided whites in those states maintaining separate institutions for the two groups. These cases met with some success in terms of improving specific segregated facilities. However, in the 1940s the "separate but equal" theory was rejected by Negro leaders (and gradually by the courts), on the ground that, in practice, segregated facilities never actually provide equal conditions and opportunities. Petitions were entered for Negro plaintiffs on the theory that exclusion from publicly operated institutions and facilities on account of race is a denial of the citizenship rights that had been guaranteed by the fourteenth amendment to the constitution. In 1954, in the case of *Brown v. Board of Education of Topeka* (347 U.S. 483), the Supreme Court held that admission to institutions of public education could not be denied on account of race and ordered those states maintaining separate educational systems to proceed with "all deliberate speed" to desegregate them. This decision was a landmark in the American Negro's struggle for equality; in effect, it nullified the position taken in *Plessy v. Ferguson* and provided the precedent for public policy in many areas other than education.

By court decisions and executive orders, the public policy of the United States has outlawed segregation in interstate travel on common carriers, the armed forces, public educational institutions, and places of public accommodation. The Negro's right to participate in the formerly "all-white" primary elections in the southern states was affirmed in 1944. Although these court decisions and orders have created new norms, they have not eliminated the fundamental disabilities from which the

masses of Negroes suffer. Counterdevices such as intimidation and violence are still effectively employed to deny many citizenship rights to Negroes.

In what has been termed the "Negro Revolution," Negroes have provided most of the effective leadership. The National Association for the Advancement of Colored People, the leading Negro organization in the struggle for Negro civil rights over the past half century, was at first joined by such newly formed groups as the Congress of Racial Equality, the Southern Christian Leadership Conference, and the Student Nonviolent Coordinating Committee. Although predominantly Negro in composition, these groups had white membership and were the beneficiaries of financial support from the white community. A corps of young Negro leaders emerged, the most celebrated of whom was the Reverend Martin Luther King—a Baptist minister and leader of the Southern Christian Leadership Conference, who won the Nobel Peace Prize in 1964. The principle of nonviolent resistance espoused by Dr. King, and by the Negro movement in general, owed much to the teachings of Mohandas Gandhi. From 1958, nonviolent demonstrations were held in numerous American cities by various groups for the purpose of presenting the enemy with a moral confrontation over the injustices suffered by Negroes. These demonstrations resulted in imprisonment for thousands of Negroes and their white allies and captured the attention of the nation.

So long as the civil rights movement was focused upon the elimination of racial discrimination in public accommodations and differential participation in programs supported by government funds, the organizations spearheading the movement were highly unified. This was expressed through frequent meetings of the Civil Rights Leadership Conference in which heads of the major organizations participated. Judicial proceedings and direct action, moreover, were appropriate techniques for the achievement of the stated objectives, and there was a consensus regarding their use. This period of unity can be said to have lasted at least through 1963, and perhaps through 1964.

As greater attention became focused on the elimination of discrimination in housing and employment and the further implementation of school desegregation and voting rights, the movement lost some of its former cohesiveness, and new ideologies and tactics began to be advocated by some of the groups. The doctrine of "black power" was advanced by the Student Nonviolent Coordinating Committee and supported by the Congress of

Racial Equality. This doctrine called for a mobilization of the economic and political power of the Negro community as solutions to the Negro's problems and as a way of providing a sense of racial identity and self-respect. It repudiated the idea that white support of and inclusion as members of Negro organizations were essential to success and viewed past alliances with whites as having contributed to the Negro's present status. Although the doctrine did not repudiate nonviolence as a technique, it asserted that violence was justified in some situations.

In 1967 the foremost exponent of "black power" was Stokely Carmichael, a young Negro college graduate and the leader of the Student Nonviolent Coordinating Committee. Carmichael pointed to the employment of Negroes as soldiers in the Vietnam war as further evidence of the general maltreatment of Negroes by the "white establishment"; while Negroes suffered discrimination at home, they were being used in disproportionate numbers in a war in which they could have no interest.

This more militant position of the Student Nonviolent Coordinating Committee and the Congress of Racial Equality was rejected by the NAACP and the Urban League, the older and larger Negro civil rights organizations, which, throughout their history, have been allied with liberal elements of the white community. For these groups, "black power" was a racist and separatist doctrine. They regarded it as an unworkable principle that was not calculated to further the cause of Negro rights. The doctrine also ran counter to the nonviolent ideology of the Southern Christian Leadership Conference, the other major civil rights organization.

In addition to impairing the unity of the civil rights organizations, the "black power" doctrine had resulted, by 1967, in a reduction of the financial support furnished the civil rights movement by whites, with the heaviest losses experienced by the more militant groups. The doctrine, moreover, had provided the rationale for open resistance by reactionary whites in what was described at the time as the "white backlash." This upsurge of white prejudice as a counterforce to Negro militancy was believed to have contributed to the success of reactionary politicians in the elections of 1966.

The federal government has played a significant role in advancing the cause of Negro rights, not only through court decisions and executive orders but also through legislation. In 1957 Congress passed the first civil rights legislation since 1875; this act, among other things, established the Civil



Rights Commission to investigate and make recommendations on racial conditions. This was followed by the civil rights laws of 1960 and 1964. The 1964 act, which is the broadest in its coverage, has titles referring to voting rights, public accommodations, public facilities, public education, and employment opportunities. Among the sanctions it authorizes, upon the finding of discriminatory conduct, is the withholding from state and local districts of funds for federally assisted programs. Its provisions on voting have been reinforced and expanded by the Voting Rights Act of 1965.

The efforts of both Negro groups and the federal government have been affected by world conditions and world opinion, as well as by the internal changes affecting minorities. The position of the Negro in the United States is obviously inconsistent with the ideologies of freedom and equality which characterize the postwar world. Furthermore, it is embarrassing to the United States government to participate in United Nations discussions and actions on colonial and minority group problems as long as discrimination is practiced within its own boundaries. Since the United Nations now includes a large number of representatives of colored peoples, the United States has become increasingly sensitive to their opinions of its racial problems. The riots occurring in American cities are known in all parts of the world and are used for propaganda purposes in the cold war.

**Great Britain.** Although the major interest in race relations today is in Africa and the United States, some attention is directed to the course of race relations in Britain, where, in the postwar years, large numbers of East Indians, Pakistanis, and West Indians have settled in English cities. With the increasing density of these groups, prejudice and discrimination have been practiced against them, and riots have occurred (Glass 1960). The reaction of the English toward these newcomers is not greatly different from that manifested by whites to the migration of Negroes to northern cities in the United States at the end of World War I, and the British government has been forced to pass legislation limiting the number and type of Commonwealth immigrants.

Race relations in the modern world are based upon color as well as upon social and cultural differences. The actions of former colonial peoples in the Near East and Middle East can be interpreted as a revolt of "colored" people against white domination. In a similar sense, the conflict between the Chinese communists and the Western powers involves color differences between the antagonists as

well as variations in ideology. The Asian-African conferences, the first of which was held at Bandung, Indonesia, in 1955, have brought together a wide variety of people who share a common consciousness of being different in color from their former rulers. Thus the question of race relations has become increasingly international, and it is no longer possible for governments to claim that their racial problems are of purely domestic concern.

G. FRANKLIN EDWARDS

[Directly related are the entries ASSIMILATION; CONSTITUTIONAL LAW, article on CIVIL RIGHTS; MINORITIES; PREJUDICE; SEGREGATION. Other relevant material may be found in COLONIALISM; RACE; SLAVERY; and in the biographies of DuBois; FANON; FRAZIER; PARK.]

#### BIBLIOGRAPHY

- ADAMS, ROMANZO 1937 *Interracial Marriage in Hawaii*. New York: Macmillan.
- BALANDIER, GEORGES 1956 *Race Relations in West and Central Africa*. Pages 145-166 in *Conference on Race Relations in World Perspective*, Honolulu, 1954, *Race Relations in World Perspective: Papers*. . . . Edited by Andrew Lind. Honolulu: Univ. of Hawaii Press.
- BERRY, BREWTON (1951) 1965 *Race and Ethnic Relations*. 3d ed. Boston: Houghton Mifflin. → First published as *Race Relations: The Interaction of Ethnic and Racial Groups*.
- BUNTING, BRIAN (1964) 1965 *The Rise of the South African Reich*. New York: Smith.
- CARR-SAUNDERS, ALEXANDER M. (1936) 1965 *World Population: Past Growth and Present Trends*. New York: Barnes & Noble.
- CONFERENCE ON RACE RELATIONS IN WORLD PERSPECTIVE, HONOLULU, 1954 1956 *Race Relations in World Perspective: Papers*. . . . Edited by Andrew Lind. Honolulu: Univ. of Hawaii Press.
- DAVISON, R. B. 1964 *Commonwealth Immigrants*. Oxford Univ. Press.
- DOVER, CEDRIC 1939 *Half Caste*. London: Secker & Warburg.
- EDWARDS, G. FRANKLIN 1959 *The Negro Professional Class*. Glencoe, Ill.: Free Press.
- FRAZIER, E. FRANKLIN (1957) 1965 *Race and Culture Contacts in the Modern World*. Boston: Beacon.
- FREYRE, GILBERTO (1933) 1956 *The Masters and the Slaves (Casa-grande & senzala): A Study in the Development of Brazilian Civilization*. 2d English language ed., rev. Translated from the 4th definitive Brazilian edition. New York: Knopf. → First published in Portuguese. An abridged edition was published in 1964 by Knopf.
- GLASS, RUTH (1960) 1961 *London's Newcomers: The West Indian Migrants*. Centre for Urban Studies, University College, London, Report No. 1. Cambridge, Mass.: Harvard Univ. Press. → First published as *Newcomers: The West Indians in London*.
- HELLMANN, ELLEN (editor) 1949 *Handbook on Race Relations in South Africa*. Cape Town and London: Oxford Univ. Press.
- HUGHES, EVERETT C.; and HUGHES, HELEN M. 1952 *Where Peoples Meet: Racial and Ethnic Frontiers*. Glencoe, Ill.: Free Press.

- KUPER, LEO 1965 *An African Bourgeoisie: Race, Class, and Politics in South Africa*. New Haven: Yale Univ. Press. → A paperback edition was published in 1965.
- KUPER, LEO; WATTS, HILSTAN; and DAVIES, RONALD L. 1958 *Durban: A Study in Racial Ecology*. London: Cape; New York: Columbia Univ. Press.
- LIND, ANDREW 1938 *An Island Community: Ecological Succession in Hawaii*. Univ. of Chicago Press.
- LITTLE, KENNETH 1948 *Negroes in Britain: A Study of Race Relations in English Society*. London: Routledge.
- LOCKE, ALAIN; and STERN, B. J. (1942) 1949 *When Peoples Meet*. New York: Hinds.
- MALINOWSKI, BRONISLAW (1945) 1961 *The Dynamics of Culture Change: An Inquiry Into Race Relations in Africa*. New Haven: Yale Univ. Press.
- MASON, PHILIP 1954 *An Essay on Racial Tension*. London: Royal Institute of International Affairs.
- OLIVER, ROLAND A.; and PAGE, JOHN D. (1962) 1963 *A Short History of Africa*. New York Univ. Press.
- PALMER, EDWARD N. 1948 Culture Contacts and Population Growth. *American Journal of Sociology* 53:258-262.
- PARK, ROBERT E. (1913-1944) 1950 *Collected Papers of Robert Ezra Park*. Volume 1: Race and Culture. Glencoe, Ill.: Free Press.
- PARK, ROBERT E. (1926) 1950 The Race Relations Cycle. Pages 149-151 in Robert E. Park, *Collected Papers of Robert Ezra Park*. Volume 1: Race and Culture. Glencoe, Ill.: Free Press.
- PARK, ROBERT E. (1939) 1950 The Nature of Race Relations. Pages 81-116 in Robert E. Park, *Collected Papers of Robert Ezra Park*. Volume 1: Race and Culture. Glencoe, Ill.: Free Press.
- PARSONS, TALCOTT; and CLARK, KENNETH (editors) 1966 *The Negro American*. Boston: Houghton Mifflin.
- PIERSON, DONALD (1942) 1947 *Negroes in Brazil: A Study of Race Contact at Bahia*. Univ. of Chicago Press.
- PRICE, A. GRENFELL 1950 *White Settlers and Native Peoples: An Historical Study of Racial Contacts Between English-speaking Whites and Aboriginal Peoples in the United States, Canada, Australia and New Zealand*. Cambridge Univ. Press.
- REUTER, EDWARD B. 1918 *The Mulatto in the United States*. Boston: Badger.
- RICHMOND, ANTHONY H. 1954 *Colour Prejudice in Britain: A Study of West Indian Workers in Liverpool, 1941-1951*. London: Routledge.
- SCHAPER, ISAAC (1940) 1941 *Married Life in an African Tribe*. New York: Sheridan.
- SILBERMAN, CHARLES E. 1964 *Crisis in Black and White*. New York: Random House.
- SIMPSON, GEORGE E.; and YINGER, J. MILTON (1953) 1965 *Racial and Cultural Minorities: An Analysis of Prejudice and Discrimination*. 3d ed. New York: Harper.
- STONEQUIST, EVERETT V. (1937) 1961 *The Marginal Man: A Study in Personality and Culture Conflict*. New York: Russell.
- THOMPSON, EDGAR T. (editor) 1939 *Race Relations and the Race Problem: A Definition and an Analysis*. Durham, N.C.: Duke Univ. Press.
- THOMPSON, EDGAR T.; and HUGHES, EVERETT C. (editors) 1958 *Race: Individual and Collective Behavior*. Glencoe, Ill.: Free Press.
- VAN DEN BERGHE, PIERRE L. 1965 *South Africa: A Study in Conflict*. Middletown, Conn.: Wesleyan Univ. Press.
- WILLIAMS, ROBIN M. JR. 1964 *Strangers Next Door: Ethnic Relations in American Communities*. Englewood Cliffs, N.J.: Prentice-Hall.
- WILSON, MONICA (HUNTER) (1936) 1961 *Reaction to Conquest: Effects of Contact With Europeans on the Pondo of South Africa*. 2d ed. Oxford Univ. Press. → The first edition was published under the author's maiden name.
- WIRTH, LOUIS; and GOLDHAMER, HERBERT 1944 The Hybrid and the Problem of Miscegenation. Pages 249-369 in Otto Klineberg (editor), *Characteristics of the American Negro*. 3d ed. New York: Harper.
- YOUNG, WHITNEY M. JR. 1964 *To Be Equal*. New York: McGraw-Hill.

## II

### SOCIAL-PSYCHOLOGICAL ASPECTS

"The problem of the Twentieth Century," wrote W. E. B. DuBois prophetically in 1903, "is the problem of the color line." And, to be sure, together with related concerns it is the problem of the twentieth century. In an era of rising expectations, "the color line" is often the salient common element underlying a diversity of other major concerns—from colonialism to poverty and war.

Important as it is, however, social scientists typically view race relations (the social interactions between any two or more socially or biologically defined "races") as merely a particularly potent special case of intergroup relations (the social interactions between any two or more identifiable groups). Concepts and principles which are helpful in understanding relations between racial groups should be of value in understanding intergroup relations in general. Indeed, scientists find this contention repeatedly borne out as the study of these phenomena progresses.

Broadly speaking, all intergroup relations are conditioned by four interrelated classes of factors: historical, sociocultural, individual, and situational (adapted from Allport 1954). The first set relates to the unique *history* of the groups and especially the history of the direct contact between the groups. Thus, no one can fully appreciate the highly patterned "group relations" of the Republic of South Africa without a thorough grounding in the peculiar and complex history of the relations between the Africans, Afrikaners, Coloureds, English, and Indians who inhabit that troubled land.

The second general class of variables relates to the *sociocultural setting* of the intergroup contact. Here a myriad of cultural, economic, political, demographic, and ecological factors come into play. Often these variables can predict at a gross level the status and change in group relations over different geographical areas; for example, in the southern United States, a variety of indexes of



Negro-white conflict and harmony—from lynching to the desegregation of public facilities—can be reliably predicted from county to county by the use of such demographic variables as the percentages of Negro and urban residents in an area (Pettigrew & Cramer 1959).

*Individual* factors constitute a third broad category. Deep-seated personality dispositions toward outgroup prejudice, as documented by the research on authoritarianism (Adorno et al. 1950), are important considerations. But so, too, are less deeply rooted personality needs to conform and gain social approval, needs which also lead to outgroup rejection when it is the ingroup sanctioned "thing to do." Likewise, individual phenomenology must be considered, for intergroup behavior is often determined less by objective realities than by the perceptions of these realities. Such perceptions can be shaped by political ideologies and "racial" myths or merely by the individual's need to perceive the situation selectively.

The final category—the face-to-face intergroup situation—provides the vital structure connecting these various levels. It is here that the historical, sociocultural, and individual factors come to a focus and, conditioned by the particular characteristics of the situation, produce intergroup relations as such. Hence, this article concentrates on an analysis of the intergroup situation.

### Intergroup contact

A fundamental issue of intergroup and race relations concerns the effects of contact. Some writers maintain that increased contact between groups of markedly different values and origins will lead only to heightened conflict; others hold that increased contact between such groups will decrease prejudice and fear and lead to greater intergroup harmony. Social science evidence supports neither extreme. Increased interaction, whether of individuals or groups, intensifies and magnifies the processes already underway. Hence, more intergroup contact can result in either greater prejudice and rejection or greater respect and acceptance, depending upon the situation in which it occurs. The basic task, then, is to specify the situational conditions under which contact leads to distrust and those under which it leads to trust.

Gordon Allport (1954), in his review of the relevant research, concluded that four characteristics of the contact situation are of the utmost importance. Prejudice and conflict are lessened when the two groups possess equal status, seek common goals, are cooperatively dependent upon each other, and interact with the positive support of authori-

ties, law, or custom. Instances throughout the world of intergroup conflict where these situational conditions are not met come readily to mind: caste contact in India, Greek-Turkish contact on Cyprus, racial contact in South Africa. But more solid evidence must come from controlled research. And opportunities for this research abound in a nation such as the United States where intergroup patterns are undergoing extensive alterations.

**Research evidence.** Thus, Allport's contact principles can be seen in operation in a series of studies of newly desegregated situations. Ira N. Brophy (1945) found that white American merchant seamen tended to hold racial attitudes in direct relation to the number of voyages they had taken with equal-status Negro American seamen: the more desegregated the voyages, the more positive their attitudes. Similarly, William M. Kephart (1957) noted that white Philadelphia policemen who had personally worked with Negro policemen were far more favorable toward the further desegregation of their force than other white policemen.

Such intergroup bonds built through optimal contact situations can even withstand severe crises. For instance, while Negro and white mobs raged during the Detroit race riot of 1943, desegregated co-workers, university students, and neighbors of long standing peacefully carried on their lives side by side (Lee & Humphrey 1943).

Mention of neighborhood desegregation introduces the most solid research evidence available. Repeated studies have found that racially desegregated living in public housing developments that meet all four of Allport's contact criteria sharply reduces intergroup prejudice among both Negro and white neighbors (Deutsch & Collins 1951; Wilner et al. 1955; Works 1961). In addition, these studies demonstrate that living in segregated, but otherwise identical, housing developments structures intergroup contact in such a manner that intergroup bitterness is, if anything, enhanced. The segregated or desegregated pattern of the housing developments fixes a *social climate* that in turn influences the intergroup expectations, behavior, and attitudes of the occupants. D. M. Wilner and his associates (1955, p. 107) concluded: "Contact and the perceived social climate tend to reinforce each other when their influence operates in the same direction and to cancel each other out when their influence works in opposite directions."

In such instances historical and sociocultural factors intersect with the intergroup situation. Consider a comparable study of a multigroup neighborhood in Durban, South Africa (Russell 1961). The area afforded a rare South African example of close

contact between Coloureds and Indians, both of relatively high occupational status, and Caucasians; and the situation did enable whites to develop somewhat more positive attitudes toward their non-white neighbors. Yet the whites in this unusual situation were defensive about their intergroup contacts, some even attempting to avoid contact as much as possible. Moreover, the contact that did take place was not truly equal and reciprocal. Whites received neighborly aid from nonwhites and entered nonwhite homes far more often than the other way around. In fact, many whites rationalized their intergroup contact in terms of the mildly exploitative aspects of the relationship. Both the whites and nonwhites were fully aware that the social climate of South Africa does not support, in fact punishes, equal-status contact between races—an atmosphere which poisons true neighborliness at the core.

A further qualification attends attitude and behavior change through intergroup contact: at least in the early stages, the change is frequently limited to the specific situation involved. Thus, J. D. Lohman and D. C. Reitzes (1952) found that white steelworkers in the northern United States generally approved of the racial desegregation of their union to the point of sharing all union facilities with Negroes and electing Negroes to high office, yet they also sternly opposed the desegregation of their all-white neighborhoods. In this case, as in many others, *institutional structures* limited the contact effects. Intergroup attitudes and behavior had changed in the work situation, with the support of the union organization; but these changes did not generalize to the neighborhood situation, where a community organization resisted desegregation.

In sum, intergroup contact can result in either conflict or harmony, depending upon the situational conditions under which it occurs. When the contact situation involves a combination of equal status, common goals, no competition along group lines, and the support of authorities, law, or custom, improved intergroup relations are likely to result. When these conditions are not met, one of two sharply contrasting negative group stereotypes often emerges.

### Two contrasting negative stereotypes

Bruno Bettelheim and Morris Janowitz (1950) have drawn a psychoanalytic distinction between outgroup stereotypes which focus upon superego concerns and those which focus upon id concerns. The superego stereotype ascribes to an entire group such personal traits as being overly am-

bitious, striving, crafty, clannish, shrewd, intelligent, sly, and dishonest. By contrast, the id stereotype ascribes to an entire group such personal traits as being overly unambitious, lazy, happy-go-lucky, loud, irresponsible, stupid, dirty, odoriferous, uninhibited, and hypersexed. The psychoanalytic interpretation of these distinctive, though strangely reciprocal, stereotypes is straightforward. These two types of outgroups serve in part as alter egos for the bigot; intergroup animosity in these terms is interpreted as a projection of the bigot's own unacceptable inner impulses onto an outgroup. The prejudiced person thus personifies his own superego sins of ambition, deceit, and egotism in the first type of outgroup, and his own id sins of the flesh in the second type of outgroup (Adorno et al. 1950; Bettelheim & Janowitz 1950).

The selection of which stereotype is to be applied to various groups is determined by a complex of historical and sociocultural factors intersecting with individual factors once again at the situational level of analysis. The superego stereotype is typically applied throughout the world to "middleman minorities," that is, groups consisting largely of merchants who are structurally squeezed between the landed and the laboring classes. The id stereotype, however, is typically invoked throughout the world to apply to groups found at the bottom of the social structure.

**The superego stereotype.** The classic middleman situation was that of the European Jews in the Middle Ages. Barred from owning land and encouraged to handle financial matters, they assumed their intermediate status because of the special historical circumstances of the period. This story has been repeated throughout recorded history, partly because the outgroup member may actually perform better than the ingroup member in market situations requiring objectivity and impersonality; *Gemeinschaft* bonds, in other words, may hinder *Gesellschaft* relations.

Cultural and racial minorities generally occupy this middleman role, probably because they serve as ideal scapegoats, as shock absorbers that protect the elite from the masses. The history of the European Jew bears tragic witness to this buffer function. During episodic crises—from depressions to revolutions and epidemics—the Jew has served as a unifying target of temporary coalitions between landed and laboring elements. Even in relatively peaceful times the Jew has often served the elite in such invidious roles as tax collector, or he has been the object of minor frustrations and competitiveness in market dealings. Hence, the superego stereotype of the Jew as ambitious, sly,



and dishonest took root. Significant, too, is the fact that this stereotype has also been applied to non-Jewish groups caught in the same middleman position. Consequently, the Chinese merchants of Malaysia and Indonesia are often called the "Jews of Asia," and the Muslim Indian merchants of east and south Africa the "Jews of Africa."

**The id stereotype.** The assignment of an impulsive id stereotype to an outgroup occurs in a wider range of intergroup situations; however, these groups are almost invariably at the bottom of the social structure and are generally far less technologically skilled than their detractors. With a history of slavery, segregation, and poverty, Negro Americans have long labored under the image of a lazy, happy-go-lucky, stupid, dirty, hypersexed people. So have the gypsies of central Europe, for centuries an outcaste group. So, too, have the poor of southern Italy. Research reveals that the view held by northern Italians of their fellow countrymen in the south is strikingly similar to the view of Negro Americans that is held by many white Americans (Battacchi 1959).

A recent example of the id stereotype has developed in Israel. Higher-status Israelis of European origins often look down on the lower-status Oriental Israelis from north Africa and Asia. The prevailing stereotype of the less technologically advanced Orientals is a familiar one. The assumed group traits are "... instability, emotionalism, impulsiveness, unreliability and incompetence . . . habitual lying and cheating, laziness, boastfulness, inclination to violence, uncontrolled temper, superstitiousness, childishness, lack of cleanliness, and in general 'primitivity' and 'lack of culture'" (Patai 1953, p. 314).

**Contradictory stereotypes.** Sometimes the two types of images are fused into a single contradictory stereotype. In Nazi Germany, for instance, the lack of a significant id type outgroup forced the anti-Semitic image of the Jew to do double duty. Jews had to serve as the personification of both superego and id concerns and were seen as both lazy and too striving, stupid and too intelligent. In most countries, however, bigots possess the luxury of a variety of outgroups, and more differentiated stereotypes have evolved. The prevalence of such stereotypes in the world today raises questions concerning the possibility of relatively permanent safeguards against intergroup conflict.

### Institutional protections

Intergroup and race relations can always be improved; group prejudice and discrimination are by no means inevitable aspects of the human con-

dition. Institutional mechanisms can be established to protect minorities and foster intergroup harmony, and often these institutional changes are won by the minorities themselves. Thus, in the United States, Negro Americans have achieved significant alterations through direct, nonviolent action (Pettigrew 1964). These alterations typically eliminate practices of racial segregation, a prerequisite for allowing and encouraging intergroup contact that meets Allport's four criteria for positive change. Indeed, nationwide public opinion polls have noted major improvements in racial attitudes and linked much of this improvement to the alleviation of segregation (Hyman & Sheatsley 1964).

**Political coalitions.** Political pressures for such institutional changes frequently come about in democracies when a coalition of minorities becomes electorally powerful. For example, the dominant political party in the United States won seven out of nine presidential elections from 1932 through 1964 as a delicately balanced coalition composed of the predominant share of the nation's Negroes, Jews, and Roman Catholics. It maintained this coalition in large part because of its achievements in establishing institutional protections of minority rights—the most heralded, perhaps, being the 1964 Civil Rights Act.

**Legally required changes.** Appropriate and enforced laws are a principal means of institutionalizing minority acceptance. Yet some repeat the nineteenth-century dictum that "laws cannot change the hearts and minds of men." Changes in race relations in the United States over the past century do not bear this out. A case in point is the 1945 legislation against discrimination in employment enacted by New York State. As a result of this law, many New York City department stores hired Negro Americans as sales clerks for the first time. Two investigations conducted separate tests of the effects of this law-induced desegregation. One study of white sales personnel revealed that those who had experienced the new equal-status job contact with Negroes held more favorable attitudes toward interracial interaction in the work situation (Harding & Hogrefe 1952). Once again, however, the initial effects of this contact did not extend beyond the immediate situation; equal-status clerks were not more likely to accept Negroes in eating and residential situations.

The second New York City department store study questioned customers (Saenger & Gilbert 1950). Though reactions to later questioning varied, there was a widespread acceptance of this legally required racial change. Customers were largely

concerned with convenient and efficient shopping; many hesitated to challenge the *fait accompli* firmly established by the law; and for others the new situation was consistent with the national belief in the justice of equal opportunity for all.

**Behavior and attitudes.** Contrary to the old adage, then, law *can* change the hearts and minds of men. It does so through a vital intermediate step. Enforced law first acts to modify behavior, and this modified behavior in turn changes men's hearts and minds. Notice that this is precisely the opposite sequence commonly believed to be the most effective method of improving intergroup relations. Persuade people to be less prejudiced through informational and good will campaigns, goes the conventional reasoning, and then their intergroup behavior will improve. To be sure, this sequence is sometimes effective, but the preponderance of social psychological evidence attests to the greater efficacy of the opposite approach. Behaving differently is more often the precursor of thinking differently.

The basic problem with exhortation as a means of bettering intergroup relations is that it does not require participants to change their behavior. If there is intergroup contact, it is generally brief and artificial. Moreover, a vast body of psychological data indicates that prejudiced individuals avoid tolerance messages altogether, deny the relevance of the messages for themselves, or find ways of twisting the meaning of the messages (Cooper & Jahoda 1947; Hyman & Sheatsley 1947).

True, antidiscrimination laws can also be "tuned out" by prejudiced individuals. But if properly enforced, such legislation has a potential for achieving behavioral change that is not possessed by exhortation. Several reasons for this are apparent. There is, of course, the threat of punishment and unfavorable publicity for recalcitrants. But more important is the "off the hook" function such laws provide. Thus, the department stores in New York City may each have been afraid to hire Negro sales personnel as long as there was no assurance that their competitors would follow suit. But the anti-discrimination law, applied to all stores, furnished this needed assurance. And, as noted in the studies previously cited, the legally established *fait accompli*, unlike exhortations for intergroup tolerance, generates its own acceptance. The situational face lifting it achieves is a year-round process, a constant, institutionalized reminder that intergroup harmony is the sanctioned norm. Finally, the new behavior required by the law commits the individual psychologically; for he who has publicly behaved in a new manner and been rewarded for

doing so is likely to become personally committed to the intergroup change.

### A microcosmic case in point

The fundamental principles underlying situational factors in intergroup relations are dramatically highlighted by an ingenious field experiment conducted by the eminent social psychologist Sherif (Oklahoma, University of . . . 1961). Twenty young boys, of homogeneous backgrounds but previously unacquainted with one another, attended a summer camp set up for the investigation. From the start, the boys were divided into two groups—"the Rattlers" and "the Eagles." The first stage of the experiment was designed to develop high *esprit de corps* within each of the groups. Totally separated from each other, the Rattlers and the Eagles engaged in a variety of satisfying experiences, and each group soon developed the pride and sense of "we-ness" characteristic of strong ingroup solidarity.

The second stage of the study brought the two groups face-to-face in a series of grimly competitive tasks, such as tug-of-war, baseball and football games, and tent pitching. In all of these contact situations only one group could win and the other had to lose. The inevitable animosity soon appeared. Derogatory songs and slogans were composed; destructive raids on "the enemy's cabin" began; negative stereotypes developed; and even preferences for group segregation were voiced. Competitive contact had wreaked its usual havoc.

The experiment's third stage tried to mend the damage. An approach involving more intergroup contact without competition came first. The boys met in such situations as eating good food in the same room, shooting off fireworks, and attending a motion picture. Note that all of these activities involved passive conduct without common goals or group interdependence. And, understandably, intergroup friction between the Rattlers and the Eagles did not abate; in fact, the boys employed these unfocused events as opportunities for further vilification of their hated rivals.

Next the investigators introduced carefully contrived problems that required the *cooperation* of both groups for their solution. Fixing the damaged water tank that supplied the entire camp, jointly raising the funds to show a favorite motion picture, and other such functionally dependent behavior achieved a striking decline in intergroup hostility. While at the close of the competitive second stage more than half of the characteristics assigned by the boys to their group rivals were sharply unfavorable, more than two-thirds of such judgments were favorable at the close of the



interdependent third stage. In addition, the percentage of friendship choices across group lines multiplied fourfold.

One final result of Sherif's intriguing investigation replicated the limited nature of the attitude change initially induced even under optimal intergroup contact conditions. The first interdependent encounters of the two camp groups by no means removed all of the bad feeling between the Rattlers and the Eagles. But as these socially sanctioned confrontations continued, the prejudice-reducing power of this type of contact accumulated.

This finding suggests that as a widespread process of institutionalized change in intergroup relations proceeds, it may well receive increasingly greater support and have increasingly greater effects upon the participants. This possibility is of special relevance to such intergroup processes as India's efforts to end caste discrimination and the United States' efforts to end racial segregation.

THOMAS F. PETTIGREW

#### BIBLIOGRAPHY

- ADORNO, THEODOR W. et al. 1950 *The Authoritarian Personality*. American Jewish Committee, Social Studies Series, No. 3. New York: Harper.
- ALLPORT, GORDON W. 1954 *The Nature of Prejudice*. Reading, Mass.: Addison-Wesley. → An abridged paperback edition was published in 1958 by Doubleday.
- BATTACCHI, MARCO W. 1959 *Meridionali e settentrionali nella struttura del pregiudizio ethnico in Italia*. Bologna (Italy): Società Editrice Il Mulino.
- BETTELHEIM, BRUNO; and JANOWITZ, MORRIS (1950) 1964 *Social Change and Prejudice, Including Dynamics of Prejudice*. New York: Free Press. → A reprinting of *Dynamics of Prejudice*, with a reassessment of its findings.
- BROPHY, IRA N. 1945 *The Luxury of Anti-Negro Prejudice*. *Public Opinion Quarterly* 9:456-466.
- COOPER, EUNICE; and JAHODA, MARIE 1947 *The Evasion of Propaganda: How Prejudiced People Respond to Anti-prejudice Propaganda*. *Journal of Psychology* 23:15-25.
- DAVIS, JOHN P. (editor) 1966 *The American Negro Reference Book*. Englewood Cliffs, N.J.: Prentice-Hall.
- DEUTSCH, MORTON; and COLLINS, MARY E. 1951 *Interracial Housing: A Psychological Evaluation of a Social Experiment*. Minneapolis: Univ. of Minnesota Press.
- DUBOIS, W. E. B. (1903) 1963 *The Souls of Black Folk: Essays and Sketches*. Gloucester, Mass.: Smith.
- HARDING, J. and HOGREFFE, R. 1952 *Attitudes of White Department Store Employees Toward Negro Co-workers*. *Journal of Social Issues* 8, no. 1:18-28.
- HYMAN, HERBERT H.; and SHEATSLEY, PAUL B. 1947 *Some Reasons Why Information Campaigns Fail*. *Public Opinion Quarterly* 11:412-423.
- HYMAN, HERBERT H.; and SHEATSLEY, PAUL B. 1964 *Attitudes Toward Desegregation*. *Scientific American* 211, no. 1:16-23.
- KEPHART, WILLIAM M. 1957 *Racial Factors and Urban Law Enforcement*. Philadelphia: Univ. of Pennsylvania Press.
- LEE, ALFRED M.; and HUMPHREY, NORMAN D. 1943 *Race Riot*. New York: Dryden.
- LOEHMAN, J. D.; and REITZES, D. C. 1952 *A Note on Race Relations in Mass Society*. *American Journal of Sociology* 58, no. 3:240-246.
- OKLAHOMA, UNIVERSITY OF, INSTITUTE OF GROUP RELATIONS 1961 *Intergroup Conflict and Cooperation: The Robbers Cave Experiment*, by Muzafer Sherif et al. Norman, Okla.: University Book Exchange.
- PATAI, RAPHAEL 1953 *Israel Between East and West: A Study in Human Relations*. Philadelphia: Jewish Publication Society of America.
- PETTIGREW, THOMAS F. 1964 *A Profile of the Negro American*. Princeton, N.J.: Van Nostrand.
- PETTIGREW, THOMAS F.; and CRAMER, M. RICHARD 1959 *The Demography of Desegregation*. *Journal of Social Issues* 15, no. 4:61-71.
- RUSSELL, MARCO 1961 *A Study of a South African Interracial Neighborhood*. Master's thesis, Univ. of Natal.
- SAENGER, GERHART; and GILBERT, EMILY 1950 *Customer Reactions to the Integration of Negro Sales Personnel*. *International Journal of Opinion and Attitude Research* 4:57-76.
- WILNER, DANIEL M.; WALKLEY, ROSABELLE P.; and COOK, STUART W. 1955 *Human Relations in Interracial Housing: A Study of the Contact Hypothesis*. Minneapolis: Univ. of Minnesota Press.
- WORKS, ERNEST 1961 *The Prejudice-Interaction Hypothesis From the Point of View of the Negro Minority Group*. *American Journal of Sociology* 67:47-52.

#### RADBRUCH, GUSTAV

Gustav Radbruch (1878-1949), German philosopher of law and statesman, was born into a well-to-do middle-class family. A baptized Lutheran, Radbruch was for many years critical of his church, but later in life he became increasingly religious, especially when confronted with the Nazi dictatorship. He then practiced a "Christianity-in-substance," professing "the belief of the unbeliever."

His excellent education in Latin and Greek at the Lübeck Gymnasium was to make him a friend of the classics for the rest of his life. His interest in social questions dated from an early age, and he began his studies of law and economics while at the University of Munich. He continued his studies in law at the University of Leipzig, where he read the works of Rudolf von Jhering and Rudolf Sohm, famous teachers at that university. At the same time, he was influenced by Lujo Brentano's idea of a liberally tempered socialism.

Radbruch attended Binding's lectures at Leipzig, and although he was less impressed by Binding's conservative opinions about the law than he was later by the new ideas that Franz von Liszt was teaching at Berlin, he nevertheless learned from Binding to do careful legal research according to the best historical methods of contemporary juris-

prudence. From boyhood Radbruch was interested in legal history, with a special liking for memoirs and anecdotes. His legal research was influenced by Karl Bücher's empirical psychologism, by Wilhelm Wundt's psychological ethnology, and by Karl Lamprecht's sociological history. Other sources of intellectual excitement were certain chapters of Marx's *Capital*, Gerhart Hauptmann's revolutionary play *The Weavers*, and Richard Dehmel's *Song of a Workman*. After he had obtained his doctorate at Berlin, Radbruch became, on Liszt's recommendation, an unsalaried lecturer at the University of Heidelberg. There he became a friend of the philosophers Emil Lask and Heinrich Levy, who brought him into contact with the "Heidelberg Neo-Kantianism" of Wilhelm Windelband and Heinrich Rickert. Inspired by this philosophy, Radbruch turned his attention from Liszt's empirical and psychological monism to the problems of a philosophy of values.

The results of these various studies were two books, *Einführung in die Rechtswissenschaft* (1910) and *Grundzüge der Rechtsphilosophie* (1914). Influenced also by Max Weber, Radbruch developed in these books two of the principles of his thought: the difference between "is" and "ought" (what he called "methodological dualism") and the axiom of "relativism" (the idea that values cannot be proved but only confessed). He then tried to combine these philosophical ideas with political theory; relativism, he asserted, stipulates tolerance, which is an indispensable condition of democracy. This thesis led Radbruch to formulate a threefold theory of political ideologies: "individualism" as applied to society produces a laissez-faire economy and gives pre-eminence to individual liberty; from "collectivism" follows the realization of a socialist economy and a general emphasis on the community; "transpersonalism"—in political practice "conservatism"—places a primary emphasis on cultural institutions.

A highly skillful teacher, Radbruch influenced a considerable number of talented foreign students, especially Polish and Russian emigrants. Together with his friend Hermann Kantorowicz, the initiator of the free-law movement, Radbruch pleaded for a sociological, functional jurisprudence (*Interessenjurisprudenz*, jurisprudence of interests) instead of the formalistic approach of the still prevailing *Begriffsjurisprudenz* (jurisprudence of concepts, analytical positivism), which conceived of the law as a body of rules applicable to every possible case regardless of the interests and social problems involved. Determined to prove his sociological ideas, he held positions in Bruchsal as town

councilor, as a member of the welfare office, and also for a time in the penitentiary.

In 1914 Radbruch became a professor of criminal law at Königsberg. When World War I began, he initially did voluntary work in a military hospital, but toward the end of the war he took part, as an officer, in several battles in France. Although a pacifist, he came to see military service as a moral obligation. "I go to prove my soul," he wrote to a friend, quoting from Browning, and added, "I want to hold solidarity with the simple folk."

When the war was over, Radbruch became a professor at Kiel. He established a new institute for adult education there (a *Volkshochschule*) and wrote articles about the need for civics instruction in the schools. During the revolt of the German army (the *Kapp Putsch*) in March 1920, Radbruch was arrested for a brief period and threatened with a death sentence, but his courage and ingenuity saved him. Shortly thereafter, he was elected to the German Reichstag and twice held the position of minister of justice; he was particularly influential in the area of penal reform. He resigned from his political positions in 1923, returning to his chair of criminal and legal philosophy at Kiel, and after 1926 he taught at Heidelberg. The subject of his inaugural lecture at Heidelberg was the challenge of political self-education (1927), and on Constitution Day, 1928, he spoke before the Reichstag on the duties of the responsible citizen in a democracy (1929).

The 1932 edition of his *Rechtsphilosophie* contained a new theory of social democracy. The principal thesis expounded was that of the unresolvable antinomies of the fundamental aspects of the idea of law: justice, legal certainty, and expediency. These "jointly govern law in all its aspects, although they may sharply contradict one another" ([1914] 1950, p. 111). Justice demands conformity to fundamental moral values; expediency (*Zweckmässigkeit*) pays attention to the social function of law; legal certainty primarily demands formal acknowledgment of the law and of judicial decisions regardless of whether or not they are just. By stressing equally each of these components, Radbruch tried to fight the fatal one-sidedness that raises any one of these principles to sole dominion. On the eve of totalitarian monism in Germany, he pleaded for the many-sidedness of the idea of law.

In his effort to "spiritualize" socialism—in accordance with Goethe's pattern for an ideal community in *Wilhelm Meisters Wanderjahre*—Radbruch aimed at a closer relationship between socialism and religion. In 1919 he had published,



together with Paul Tillich, the *Religionsphilosophie der Kultur*, and in 1922 he wrote the *Kulturlehre des Sozialismus*, with the intention of reconciling socialism with humanism.

In April 1933, the Nazi regime dismissed Radbruch as an "untrustworthy official." His house was searched and his private papers confiscated. Thus excluded from public activity, he began to write a biography of the famous criminologist, Anselm von Feuerbach. This important work, highly esteemed by Thomas Mann, was published in Vienna in 1934. Another remarkable book, *Elegantiae juris criminalis* (1938) contains essays on historical subjects. *Gestalten und Gedanken* (1944), a collection of beautiful studies on art and literature, was written during the war.

When Radbruch was reinstated in 1945, he was 67 years old and no longer in good health. His only son had died during the war in Russia. Unable to resume the burden of political activity, he contented himself with giving lectures. He did have the consolation that during his compulsory 12-year absence from the German academic scene, his works were greatly appreciated abroad. In 1935 he had spent several months visiting University College, Oxford, and studying English law. These studies resulted in *Der Geist des englischen Rechts* (1946a) and in several essays, of which "Gesetzliches Unrecht und übergesetzliches Recht" (1946b) became the most famous. This essay shows a tendency toward natural-law doctrines, an inclination to evolution, rather than revolution, in Radbruch's views. On his seventieth birthday, on November 11, 1948, a *Festgabe* entitled *Beiträge zur Kultur- und Rechtsphilosophie* was dedicated to him (*Beiträge* . . . 1948). One year later he had a stroke and died.

The works of Radbruch were frequently criticized during his lifetime by German writers who found fault with his "materialistic positivism," his "enlightened rationalism," and his "unscientific relativism." Conservatives blamed him for being revolutionary, liberals for being a socialist, communists for being bourgeois. Idealists called him a realist, historians labeled him a philosopher, and philosophers rebuked his "simple jurisprudence." Sociologists disliked his "poetical sense"; lawyers could not accept his "humanistic bent." The abundance of Radbruch's spirit, his tolerance, and humanity were as irritating to certain narrow-minded people as his versatility was shocking to others. Now he has won recognition, and it is realized that his relativism is of the greatest value for the necessary self-criticism of the law and of lawyers. He revealed the provisional character of all human

laws, the questionable character of all self-righteous jurisdiction, and the inhumanity of all intolerant totalitarianism. He became one of the most important figures in modern legal philosophy in Germany and acquired a receptive audience in the rest of the world.

ERIK WOLF

[For the historical context of Radbruch's work, see the biographies of BRENTANO; BÜCHER; LAMPRECHT; WEBER, MAX; WUNDT. For discussion of the subsequent development of Radbruch's ideas, see JURISPRUDENCE; LAW; POLITICAL THEORY; and the biographies of BRECHT; KANTOROWICZ; Kelsen.]

#### WORKS BY RADBRUCH

- (1910) 1961 *Einführung in die Rechtswissenschaft*. 10th rev. ed. Stuttgart (Germany): Koehler.
- (1914) 1950 Gustav Radbruch: *Legal Philosophy*. Pages 43-224 in *The Legal Philosophies of Lask, Radbruch, and Dabin*. Translated by Kurt Wilk. 20th Century Legal Philosophy Series, Vol. 4. Cambridge, Mass.: Harvard Univ. Press. → First published as *Grundzüge der Rechtsphilosophie*. The 1950 edition was translated from the revised and rewritten edition of 1932, published as *Rechtsphilosophie*. A sixth German edition, edited by Erik Wolf, was published in 1963.
- 1919 RADBRUCH, GUSTAV; and TILlich, PAUL *Religionsphilosophie der Kultur: Zwei Entwürfe*. Berlin: Reuther. → "Über Religionsphilosophie des Rechts," by Gustav Radbruch; "Über die Idee einer Theologie der Kultur," by Paul Tillich.
- (1922) 1949 *Kulturlehre des Sozialismus: Ideologische Betrachtungen*. 3d rev. ed. Berlin: Dietz.
- 1927 *Der Mensch im Recht: Heidelberger Antrittsvorlesung*. Tübingen (Germany): Mohr.
- 1929 *Verfassungsrede gehalten von prof. dr. Gustav Radbruch bei der Feier der Reichsregierung am 11. august 1928*. Berlin: Reichszentrale für Heimatdienst.
- (1934) 1957 *Paul Johann Anselm Feuerbach: Ein Juristenleben*. 2d ed., edited by Erik Wolf. Göttingen (Germany): Vandenhoeck & Ruprecht.
- 1936a Anglo-American Jurisprudence Through Continental Eyes. *Law Quarterly Review* 52:530-545.
- 1936b Justice and Equity in International Relations. Pages 1-13 in *Justice and Equity in the International Sphere*. New Commonwealth Institute Monographs, Series B, No. 1. London: Constable.
- 1936c Jurisprudence in the Criminal Law. *Journal of Comparative Legislation and International Law* 18: 212-225.
- (1938) 1950 *Elegantiae juris criminalis: Vierzehn Studien zur Geschichte des Strafrechts*. 2d ed., rev. Basel: Verlag für Recht und Gesellschaft.
- 1944 *Gestalten und Gedanken: Acht Studien*. Leipzig: Koehler & Amelang.
- (1946a) 1958 *Der Geist des englischen Rechts*. 4th ed. Göttingen (Germany): Vandenhoeck & Ruprecht.
- (1946b) 1950 *Gesetzliches Unrecht und übergesetzliches Recht*. Pages 347-357 in Gustav Radbruch, *Rechtsphilosophie*. 4th ed. Stuttgart (Germany): Koehler.

#### WORKS ABOUT RADBRUCH

*Beiträge zur Kultur- und Rechtsphilosophie*. 1948 Heidelberg (Germany): Rausch. → See especially pages 280-287 for a bibliography of Radbruch's works.

- BODENHEIMER, EDGAR 1962 *Jurisprudence: The Philosophy and Method of the Law*. Cambridge, Mass.: Harvard Univ. Press. → See especially pages 131–134.
- BRECHT, ARNOLD 1959 *Political Theory: The Foundation of Twentieth-century Political Thought*. Princeton Univ. Press. → See especially pages 233–236.
- CAMPBELL, ARCHIBALD H. 1949 *Gustav Radbruch's Rechtsphilosophie and the English Jurisprudence*. Preface by Dr. E. Wolf. Münden (Germany): Nowack.
- CHROUST, ANTON H. 1944 *Philosophy of Law of Gustav Radbruch*. *Philosophical Review* 53:23–45.
- FRIEDMANN, WOLFGANG (1944) 1960 *Legal Theory*. 4th ed. London: Stevens.
- FULLER, LON L. 1954 *American Legal Philosophy at Mid-century*. *Journal of Legal Education* 6:457–485. → See especially pages 481–485 on the legal philosophy of Gustav Radbruch.
- PATTERSON, EDWIN W. 1953 *Jurisprudence: Men and Ideas of the Law*. New York: Foundation Press. → See especially pages 399–403.
- WOLF, ERIK 1958 *Revolution or Evolution in Gustav Radbruch's Legal Philosophy*. *Natural Law Forum* 3: 1–23.

### RADCLIFFE-BROWN, A. R.

Alfred Reginald Radcliffe-Brown (1881–1955) was one of the most eminent anthropologists of the first half of the twentieth century. By example and teaching he helped to develop and establish modern “social” anthropology as a generalizing, theoretical discipline. The most notable of his many important contributions was his application to primitive societies of some of the ideas of systems theory, which led to a revolution in the analysis and interpretation of social relations. In brief, he may be said to have turned social anthropology from its preoccupation with historical development and psychological extrapolation to the comparative study of persistent and changing social structures.

Radcliffe-Brown was born and educated in England. He attended first the Royal Commercial Travellers' School at Pinner, Middlesex, and then for two years was a foundation scholar at King Edward's High School, Birmingham. He left school when not yet 18 years old, undecided on a career and with few prospects. By private study he gained an exhibition at Trinity College, Cambridge, in 1901 and held it with increasing distinction until 1906. His original intention had been to study natural science, but his tutor, W. W. Rowse Ball, a mathematician and minor historian, and incidentally a great admirer of Sir James Frazer, diverted him to “moral science.” In the curriculum of the time, moral science included experimental psychology and economics as well as philosophical subjects. Radcliffe-Brown took his bachelor's degree in 1905, being placed in the first division of the first class.

Of great significance were his university associations with W. H. R. Rivers, A. C. Haddon, and C. S. Myers. All three had been members of the 1898 Cambridge Expedition to Torres Strait. Myers was the exemplar of rigorous and fruitful scientific method, while Haddon and Rivers introduced Radcliffe-Brown to the discipline of anthropology and vitally influenced the shaping of his approach. Haddon, who became one of Radcliffe-Brown's most enthusiastic sponsors, transferred to him his own critical acumen, his interest in classification and morphology, his demand for inductive generalization on the widest possible basis, and his recognition that a strict comparative method demands the intensive field study of particular societies. There is some reason to think it was Haddon, too, who first made him critically aware of the systemic interdependence of social phenomena and who may thus have edged him toward a sympathy for Durkheim's viewpoint. Rivers was an inspiring teacher in psychology, but Radcliffe-Brown progressively turned away from Rivers' conception of anthropology, which eventually became a historical fantasy about the diffusion of culture. However, the qualities of mind that Rivers showed in his psychological studies—his insistence on scientific procedures, delight in analysis, and facility in adapting problems to novel experimental conditions—were precisely those which Radcliffe-Brown developed to a striking degree. He dedicated *The Andaman Islanders* (1922) to both Haddon and Rivers.

Given this background, it is not difficult to understand the fixation on the methods of natural science that came to characterize Radcliffe-Brown's approach. It is noteworthy, also, that he studied carefully the writings on the philosophy of science of William Whewell, especially Whewell's work on the processes of inductive thought. In view of the recent rediscovery of Whewell's importance, it says much for Radcliffe-Brown's precocity that he preferred him to J. S. Mill at a time when Mill was generally much more highly regarded.

Most of Radcliffe-Brown's working life was spent outside England. He held chairs of social anthropology successively at Cape Town, 1920–1925; Sydney, 1925–1931; Chicago, 1931–1937; and Oxford, 1937–1946, where he was also a fellow of All Souls College. He was visiting professor at Yenching in 1935 and at São Paulo from 1942 to 1944. After his retirement from Oxford he was professor of social science and director of the Institute of Social Studies at Farouk I University, Alexandria, from 1947 to 1949, and he later held a special appointment at Rhodes University, Gra-



hamstown, South Africa, from 1951 to 1954. He was associated in various capacities with several other universities, including Cambridge (where he had been a fellow of Trinity College from 1908 to 1914), London, Birmingham, and Manchester.

He devoted a great deal of time to the stimulation and organization of research by others and was restlessly active in promoting large schemes. While in South Africa, he organized the School of African Life and Languages with his own chair as a nucleus. At Sydney, in conjunction with the Australian National Research Council, he founded the journal *Oceania* and directed a vigorous and successful research program. But there, as at Cape Town and later at Chicago, he was frustrated by a shortage of money, and he had the same problem at Oxford, where there was an interest in research but no substantial funds available until after he had retired.

Among his many professional distinctions were membership in the Amsterdam Royal Academy of Sciences, honorary membership in the New York Academy of Sciences, fellowship in the British Academy, first presidency of the British Association of Social Anthropologists, and the presidency of the Royal Anthropological Institute, which awarded him the Rivers Medal in 1938 and the Huxley Memorial Medal in 1951. Only the Australian National Research Council declined to admit him to full membership. A number of his students who attained high academic distinction in the field of anthropology, together with others whom he had influenced less directly, paid him tribute in two collections of essays (*Social Anthropology* . . . 1937; Fortes 1949) and by two special printings of some of his own outstanding writings ([1923–1949] 1961; 1958).

A tall man with a distinguished air and presence, strikingly handsome in his prime, he often captivated people as much by his charm, wit, and cultivation as by the appeal of his ideas. He was a particularly brilliant teacher who was so much the master of his subject, and of the arts of rostrum and seminar, that he could expound the most difficult topics without notes or any outward hint of preparation. His writings gave the same impression.

Throughout his life, his complex personality caused divided reactions to him, arousing devotion, ambivalence, and hostility. In his youth he had a reputation of dash, extravagance, and overbrilliance, and at Cambridge he had been known as "Anarchy" Brown because of a flirtation with anarchism, later transmuted into a mild socialism. (He changed his name by deed poll, in 1926, from

Brown to Radcliffe-Brown, Radcliffe being his mother's family name.) When more mature he tended to keep people at a certain distance and seemed to discourage overintimacy or dependence, rarely permitting himself to show signs of private experience. Although he did not seek disciples, circles did form around him. In congenial company, he was warm and kindly and, with students, he was patient, courteous, and almost always helpful. Several generations of young people thought him inspirational and found that the inspiration lasted. But there were those to whom he showed another side, who felt his cutting wit and memorable power of scorn. He could give an impression of waiting for others to cross a *pons asinorum* that he had built for them. He was somewhat given to instructing other scholars in their own subjects and his conception of social anthropology emboldened him to call into question the autonomy or status of other disciplines. Many who knew him over long periods thought him, certainly, a man of *hubris*, yet at the same time without jealousy, malice, or censoriousness. He had no liking for academic politics and was not caught up in the pursuit of power or advantage. He was never affluent, and his health caused him more or less constant concern.

**Theoretical orientation.** Radcliffe-Brown's outlook rested on a highly personal philosophy of science. A primary influence on that philosophy was Whewell, from whom he appears to have acquired his passion for method. He possessed in a notable degree what Whewell had regarded as an all-important requisite in inductive science—a fertile, sagacious, ingenious, and honest mind—and his work continually exemplified Whewell's formula for the growth of a generalizing science of principle—the "colligation of facts" and the "explication of concepts" by "progressive intuition."

In its more general philosophical aspect, his position was a synthesis, or attempted synthesis, of extraordinarily diverse elements: he drew his theory of reality from Heraclitus, his theory of process from Herbert Spencer, and his theory of epistemology from Durkheim. Moreover, his indebtedness was by no means limited to those thinkers. There is evidence that he drew on Hume, Samuel Alexander, Whitehead, and Ralph Barton Perry, and he had a considerable affinity for Chinese philosophy. He freely acknowledged the influence of certain writers of the French Enlightenment, notably Condorcet and Montesquieu, in forming his conception of social science. Comte and Durkheim had considerable attraction for him, and he spent much effort in testing some of Durkheim's ideas against

facts. His debt to these Continental theorists, however, was more for their sociology than for their philosophy, a fact that is somewhat concealed by his proximate debt to Spencer. Although Radcliffe-Brown did not reject the agnostic-evolutionist label, he had important disagreements with Spencer, just as he did with Comte and Durkheim: his most serious disagreement with Spencer had to do with the latter's historical speculations, his extreme individualism, and the utilitarianism to which Durkheim also objected. Radcliffe-Brown agreed with Durkheim about the relation between the individual and society but ruthlessly pruned the French sociologist's thought of its reified abstractions.

His conviction that the particular events of social life are the facts to which all concepts and theories must be applied rested directly on a Heraclitean view of reality. His critics never realized that his fundamental viewpoint was thoroughly historical, that he merely claimed Hume's "privilege of the sceptic" as to the possibility of making inductive generalizations about history; its entanglement of accident with law seemed to him to rule out the prospect. The Heraclitean logos, with its emphasis on persistence through change, on the necessary interconnectedness of things, and on formal rather than genetic unity, posed an essential problem: to find what is discoverably coherent in the social process. It is not *always* changing in *all* respects *at once*, but is constrained and shaped to be what it is at particular places and times. This view, in amalgam with Spencer's conception of evolution as at one and the same time a process toward higher integration and differentiation, and with some contributions from Durkheim's sociology, led Radcliffe-Brown to the conception of types and forms of social structure as fiduciary equilibria in persistent systems.

**Theoretical contributions.** Radcliffe-Brown formed his theoretical approach as early as 1908, when as a postgraduate student he stated the requirements of a science of human society. He considered them to be threefold: to treat social phenomena as natural facts and thus subject to discoverable necessary conditions and laws; to adhere to the methodology of the natural sciences; to entertain only generalizations that can be tested and verified. He never departed from these rules, although his conceptual thought developed steadily.

His ideas and methods did not change essentially from those he put forward in his early works, *The Andaman Islanders* and two papers, "The Methods of Ethnology and Social Anthropology" (1923) and "The Mother's Brother in South Africa" ([1924] 1961, pp. 15-31); but their classic formulation

came somewhat later, in two years, "On the Concept of Function in Social Science" ([1935] 1961, pp. 178-187) and "On Social Structure" ([1940] 1961, pp. 188-204), and in a third study, in which he applied his theoretical approach, *The Social Organization of Australian Tribes* (1931a).

Instead of explaining social phenomena in historical or psychological terms, which he believed to be impossible, Radcliffe-Brown proposed to explain them as persistent systems of adaptation, coaptation, and integration. His main working hypothesis was that the life of a society can be conceived of as a dynamic fiduciary system of interdependent elements, functionally consistent with one another. He had used the notion of "social structure" as early as 1914, but in Frazer's and Rivers' rather ill-defined sense, as almost a doublet of "organization." In the 1920s his use of the notion became more explicit, and in the 1930s quite precise. In his final formulation, structure refers to an arrangement of persons and organization to an arrangement of activities. At the same time, he substituted the concept of "social system" for that of "culture." All these changes were connected.

In order to achieve scientific explanation, Radcliffe-Brown urged that anthropology free itself from concern with what Whitehead called "the goading urgency of contingent happenings." Although he was himself a humanist, he saw that a humanist anthropology was premature and could prevent wide induction, comparison, and generalization. His primary goal was the abstraction of general features and the search for comparable types and varieties, and he believed that the only acceptable method for acquiring systematic knowledge is to test successive hypotheses with facts. Some of his contemporaries who admired his empirical and analytical studies nevertheless failed to appreciate the extent to which these studies derived their excellence from his methods and principles; instead, they felt that his anthropology was unduly sparse, rigid, and lacking in human values. The impression that Radcliffe-Brown's work was surrounded by an aura of unreality was created by his abstract conception of anthropology as a science that could move from empiricism, classification, and unguided induction to postulation and many-dimensional theory. It was a conception that anthropologists with a historical, genetical, or psychological outlook, including those of Malinowski's school, could not accept, and indeed his idea of social anthropology as a "comparative sociology," with the fundamental character of a natural theoretical science, did not win the recognition Radcliffe-Brown had hoped for. To be sure, many of



the empirical and analytical discoveries that he made only by virtue of that conception, as well as his general principles of functional-structural study, came into wide use, but within frameworks of thought and in the service of methods having little in common with his.

His published work was slender in bulk, comprising only some 70 items, even including miscellaneous writings such as reviews. All of his writings are marked by clear language, impeccable style, and logical ability, combined with exceptional scientific imagination. He also had a sense for good technical language and for classification and typology. He gave many useful technical notions and terms to anthropology: for example, a precise language for the orders of family and kin relationships; the distinctions between *pater* and *genitor*, between rights *in rem* and *in personam*, and between organization and structure; and such notions as that of a "corporation" serving an "estate," of "alliance" or "consolidation," of "structural opposition," and of "ritual status" and "ritual value." His schematic ability was well displayed even in his first Australian study, "Three Tribes of Western Australia" (1913).

His only extended field expeditions were to the Andaman Islands, in 1906-1908, and to north-western West Australia, in 1910-1912, but there is no substance to the allegation that he had a temperamental preference for the armchair. Rather, both the studies that resulted from his field trips and those that were produced from secondary research reveal his characteristic use of theory to guide imagination. In "The Mother's Brother in South Africa," not the product of a field expedition, he developed a brilliant hypothesis of correlative, sympathetic, and antithetic functions to account for similar structural patterns of relationships in diverse types of societies. In *The Social Organization of Australian Tribes* (1931a), which covered the whole of aboriginal Australia as then known, he made a catalogue, classification, analytic generalization, and synthesis of an immense range of data on patterns of sex and age, community of language and custom, possession and occupancy of territory, kinship, marriage, segmentation, and cosmology. From this monumental account of variety within homogeneity there emerged, in addition to a typology that guided thought for a generation, the first forms of the analytical generalizations of relationship systems on which "structure" study was to be based, notably the jural significance of kinship classifications; the structuring of relationships by generation, age, and sex; the "internal solidarity" of sibling groups; and

the "external unity" of lineage. These two studies alone would have assured him a substantial reputation, but he produced 14 others on totemism (1914; 1929), primitive law ([1933] 1961, pp. 212-219), sanctions ([1934] 1961, pp. 205-211), patrilineal and matrilineal succession ([1935] 1961, pp. 32-48), taboo ([1939] 1961, pp. 133-152), joking relationships ([1940; 1949] 1961, pp. 90-104, 105-116), religion ([1945] 1961, pp. 153-177), kinship systems ([1941] 1961, pp. 49-89), theory of comparative social anthropology (1958, pp. 42-129), and political systems (1940, pp. xi-xxiii).

In constructing his theory of social systems, Radcliffe-Brown considered "phenomenal intelligible reality" to consist of objects or events and the relations between them. The relations are of two kinds, which may be symbolized as *R* and *r*. The first kind, *R*, are spatiotemporal relations of "real interconnectedness"; the second, *r*, are logico-mathematical relations which are "immanent in the universe" and independent of space and time. He conceived of social anthropology as a discipline that ultimately would deal theoretically with both classes.

The social anthropology of function, structure, and relational networks (the first phase in the development of social anthropology more generally) deals with the relations of real interconnectedness, with "... the continuing arrangement of persons in relationships defined or controlled by institutions, i.e., socially established norms or patterns of behavior" (1958, p. 177). The substance of this study is the "real and concrete" social structure resulting from "role-activities" of persons acting from "positions" in that structure. Interrelational (*R*) concepts apply only to what he called "the internal nature" of particular social systems, such as those of Kariera or Aranda, a system being a set or assemblage of interdependent parts forming "a naturally occurring unity," a complex, ordered, and unified whole in a particular region over a period of time.

In the second phase of the development of social anthropology, an effort would be made to deal with *r*-relations, which Radcliffe-Brown conceived to be, at their simplest, relations of similarity and difference. This would require some sort of nonquantitative mathematics or other system of symbols. Although he held that *R*-relations were different from *r*-relations, it was characteristic of him to envisage an eventual theoretical science bringing both within mathematical or symbolic analysis, the task of abstract theory being the conceptualization of "real interconnectedness" in ways that would

make the analysis possible. He realized such a science existed only "in its most elementary beginnings," and he himself devoted considerable effort to the task of bringing the two classes of relationship together.

Radcliffe-Brown's social anthropology is best described by separating two main elements, a general theory and a central one. The general theory produced three connected sets of questions. The first set deals with *static*, or *morphological*, problems: what kinds of societies are there? what are their similarities and differences? how are they to be classified and compared? The second set deals with *dynamic* problems: how do societies function? how do they persist? The third deals with *developmental* problems: how do societies change their types? how do new types come into existence? what general laws relate to the changes? The general theory dealing with these problems was transposed from biology and bore a heavily Spencerian cast in its emphasis on three aspects of adaptation: ecological adaptation to the physical environment; social adaptation, i.e., the institutional arrangements by which social order is maintained; and the socialization, or "cultural adaptation," of persons.

The central theory dealt with the determinants of social relations of all kinds. Radcliffe-Brown phrased it in terms of the coaptation or fitting together or harmonization of individual interests or values that makes possible "relations of association" and "social values." The theory resembles Spencer's "market" model of interaction and draws on the tradition reflected in Ralph Barton Perry's *General Theory of Value*.

The two theories are articulated in the idea that the life of a society can be conceived and studied as a system of relations of association and that a particular social structure is an arrangement of relations in which the interests or values of different individuals and groups are coapted within fiduciary "social values" expressed as institutional norms.

The idea of coaptation is fundamental to Radcliffe-Brown's whole outlook, but the logical and conceptual implications are not fully worked out, nor are the static and dynamic aspects of the coaptive process. What he did write is probably best viewed as only a sketch for a "pure" theory dealing with all classes of relations of association and, a fortiori, all classes of functioning systems or social structures. He became increasingly preoccupied with the general theory and, in his later period especially, with the concept of structure, at the expense of the theory of coaptation.

While Radcliffe-Brown did not regard the study of social structure as the whole of anthropology, he did consider it to be its most important branch; but he asserted that "the study of social structure leads immediately to the study of interests or values as the determinant of social relations" (1940) and that a "social system can be conceived and studied as a system of values" ([1939] 1961, pp. 133-152).

W. E. H. STANNER

[For the historical context of Radcliffe-Brown's work, see the biographies of COMTE; DURKHEIM; FRAZER; HADDON; KROEBER; MALINOWSKI; RIVERS; SPENCER; WHEWELL; for discussion of his ideas, see ANTHROPOLOGY, article on SOCIAL ANTHROPOLOGY; CULTURE; ETHNOLOGY; FUNCTIONAL ANALYSIS; KINSHIP.]

#### WORKS BY RADCLIFFE-BROWN

Dates in brackets in text are dates of first publication and are included in the collection 1923-1949.

- 1913 Three Tribes of Western Australia. *Journal of the Royal Anthropological Institute of Great Britain and Ireland* 43:143-194.
- 1914 The Definition of Totemism. *Anthropos* 9:622-630.
- (1922) 1948 *The Andaman Islanders*. Glencoe, Ill.: Free Press.
- 1923 The Methods of Ethnology and Social Anthropology. *South African Journal of Science* 20:124-147.
- (1923-1949) 1961 *Structure and Function in Primitive Society: Essays and Addresses*. London: Cohen & West; New York: Free Press.
- 1929 Notes on Totemism in Eastern Australia. *Journal of the Royal Anthropological Institute of Great Britain and Ireland* 59:399-415.
- 1930a Editorial. *Oceania* 1:1-4.
- 1930b Former Numbers and Distribution of the Australian Aborigines. *Official Yearbook of the Commonwealth of Australia* 23:687-696.
- 1930c The Sociological Theory of Totemism. Pages 295-309 in Pacific Science Congress, Fourth, Batavia-Bandoeng (Java), 1929, *Proceedings*. Volume 3: Biological Papers. The Hague: Nijhoff.
- (1931a) 1948 *The Social Organization of Australian Tribes*. Glencoe, Ill.: Free Press.
- 1931b Applied Anthropology. Australian and New Zealand Association for the Advancement of Science, *Report* 20:267-280.
- 1940 Preface. In Meyer Fortes and E. E. Evans-Pritchard (editors), *African Political Systems*. Oxford Univ. Press.
- 1950 Introduction. Pages 1-85 in A. R. Radcliffe-Brown and Daryll Forde (editors), *African Systems of Kinship and Marriage*. Oxford Univ. Press.
- 1958 *Method in Social Anthropology: Selected Essays*. Edited by M. N. Srinivas. Univ. of Chicago Press.

#### SUPPLEMENTARY BIBLIOGRAPHY

- ELKIN, A. P. 1956 A. R. Radcliffe-Brown: 1881-1955. *Oceania* 26:239-251.
- FIRTH, RAYMOND 1956 Alfred Reginald Radcliffe-Brown: 1881-1955. *British Academy, London, Proceedings* 287-302.
- FORTES, MEYER (editor) 1949 *Social Structure: Studies Presented to A. R. Radcliffe-Brown*. Oxford: Clarendon.



- FORTES, MEYER 1955 Radcliffe-Brown's Contributions to the Study of Social Organization. *British Journal of Sociology* 6:16-30.
- FORTES, MEYER 1956 Alfred Reginald Radcliffe-Brown, F.B.A.; 1881-1955: A Memoir. *Man* 56:149-153.
- LOWIE, ROBERT H. 1937 *The History of Ethnological Theory*. New York: Farrar & Rinehart. → See especially pages 221-229 on "Radcliffe-Brown."
- REDFIELD, ROBERT (1937) 1955 Introduction. In *Social Anthropology of North American Tribes*. Enl. ed. Edited by Fred Eggan. Univ. of Chicago Press.
- Social Anthropology of North American Tribes*. Enl. ed. Edited by Fred Eggan. (1937) 1955 Univ. of Chicago Press.

## RADIATION

This article is devoted to examining the branch of psychology that is concerned with the behavior of living organisms after their exposure to ionizing radiations. The subject is a relatively new one in the field of psychology and is allied most closely to topics subsumed under the headings "physiological psychology" and "radiobiology." At various times the subject has fallen under the headings "radiation psychology" and "radiopsychology" or, most often, under various forms of the heading "effects of radiation upon behavior."

The field has been of practical interest to the medical sciences since the discovery of X rays in 1895, although behavioral scientists did not enter the field formally for another three decades. Since the inception of radiation research, the biomedical literature has reflected concern with behavioral data in terms of their application to diagnosis, abnormal growth patterns, heredity, and lethality. From the beginning, biomedical scientists found that behavioral effects of radiation provided substantive data to aid in the understanding of the structure and function of organ systems. Perhaps the most recent application of this field to applied sciences has taken place since the advent of the space age, when those concerned with the effects of unusual environmental conditions upon man have focused attention on the radiation hazards in space (Hanrahan & Bushnell 1960, pp. 155-188).

The influence of radiation no longer remains the exclusive interest of radiation scientists. Nuclear weapons have vastly widened the scope of attention, and the layman has long recognized that he lives in an age when radiation is a highly publicized fact of life. The employment of atomic energy for peaceful uses has led to the introduction of nuclear power plants into community life. It has now become relevant to acquaint the nonspecialist with the legal implications of radiation for labor and management. It is to the general interest of

all to understand the relationship of radiation to all living matter so that appreciation may be developed for the opportunities and the complications now presented to us. Power plants will create more and cheaper energy sources; but power plants will create health hazards and contamination. New concepts of waste disposal will have to be developed, and these must be integrated with the thinking of city planners, oceanographers, sociologists, economists, and others who are concerned with the rapidly changing modern world. Those interested in population change will find the radiation scientists have much to say about processes of aging and heredity. It will be of interest to the geopolitically minded that radiation psychologists are investigating the social dynamics of groups who have been, or who may be, exposed to radiation. To those involved in influencing public opinion, the implications of this field are self-evident in a world of ideological conflict.

The problems of studying the effects of radiation are numerous and complex (see Bibliography). The extent of this complexity is well described by Van Cleave (1963, p. 356), who stresses that most new discoveries "lead to new horizons of the unknown."

**Major variables to be studied.** The psychologist interested in radiation is concerned with relating his behavioral data to their underlying physiological processes. This requires that the following variables must be assessed from both the psychological and biological point of view, under comparable environmental conditions:

**The stimuli.** There are two classes of ionizing radiations, material and electromagnetic. The former include neutrons, alpha and beta particles, deuterons, and protons; the latter include X rays and gamma rays. All, by definition, are capable of changing cell structure by dislodgement of electrons from atoms. (The references provide extensive descriptions of the characteristics of these stimuli.) The psychologist studying radiation must concern himself with them individually and in combinations. Thus, he is faced with examining the effects of many energy levels, e.g., from 70 kilovolts to several million electron volts for X rays and gamma rays, from 1,000,000 to 14,000,000 electron volts for fast neutrons, and so on, for each of these radiations.

**Method of exposure.** Radiation is known to affect the target differentially, depending on how it is delivered. For example (Bond et al. 1965), doses of a few hundred rads to the whole body will cause death within days, while doses of several thousands of rads are delivered to focal points of the body

in radiation therapy without lethal effect. The effects of exposure, therefore, need to be considered in terms of whether the dose is delivered to the whole body or to a part of it. Both types of exposure must be related to frequency and duration of exposure and to rates of delivery of the dose over various distances from source to target. Any shielding conditions in the exposure environment must likewise be examined.

*Behavior to be studied.* All behavior is the object of the radiation psychologist's investigations. The affective, cognitive, sensory, and motor aspects of an individual organism's behavior, as well as social interaction of the radiated individuals, are considered.

*Subjects to be employed.* All living species are potential objects of investigation. Special consideration must be given to such factors as age at time of exposure, age at time of behavioral study, sex, and inherited and acquired characteristics related to the behavior to be studied. Certain animals become subjects of choice because of their known biological characteristics, the presence of which may aid in the clarification of a specific question related to a radiation effect.

*Instrumentation and procedures.* The equipment and methods employed may be those standardized by experimental psychology, or they may be especially devised to test for unique aspects of behavior elicited by radiation. Experimenters have frequently used normal control subjects for comparison with experimental populations, but they have just as often used radiated subjects as their own controls for preradiation and postradiation capabilities.

*Nature of studies undertaken.* An examination of the experimental findings in the behavioral field reveals attempts in many parts of the world to explore the aforementioned variables. The many recent and comprehensive reviews of the subject permit the reader to follow the work to a degree impossible to duplicate in this brief report (Stahl 1959; 1960; Lebedinskii et al. 1958; Russell 1954; Rugh 1962; 1963; Furchtgott 1956; 1963; 1964; Van Cleave 1963; International Symposium . . . 1964; WHO 1957; Northwestern Univ. 1962; Lebedinskii & Nakhil'nitskaia 1963; Bond et al. 1965; Kimeldorf & Hunt 1965). All of these sources call attention to the relevant allied science reviews and provide extensive bibliographies.

Kaplan (1962, pp. 153-160) has provided an additional review, in which he calls attention to the techniques and procedures employed in the field. Although this paper is limited to an analysis of the rat maze experimentation, it conveys im-

plications about the problems associated with techniques involving all species, with all behavioral subject matter. The summaries of all the referenced compilations yield the following types of information:

(1) A wide assortment of living organisms have been employed as subjects. Men, subhuman primates, dogs, cats, rodents, fowl, fish, and many invertebrate forms have variously been exposed to X, gamma, alpha, beta, and neutron radiation, as well as to mixtures of these. Dose levels have been given from fractions of a roentgen (r) to over 100,000r. Rates of radiation delivery to both the whole body and to focal points of the body vary from experiment to experiment. Some studies have been devoted to single and others to additive radiation exposures. Studies have utilized both lethal and sublethal doses.

(2) Learning, general activity, emotionality, mating behavior, classical and instrumental conditioning, and visual, auditory, vestibular, cutaneous, olfactory, and kinesthetic functions have been explored. The employment of mazes, classical and operant conditioning devices, primate general test apparatus, and open field boxes are among the many testing instruments employed.

(3) Differences in procedures and apparatus and failure of behavioral scientists to standardize methods and equipment have led to slow advancement and difficulty in generalizing from available findings.

*Effects of radiation on behavior.* While a vast amount of knowledge has been accumulated by all the disciplines concerned with radiation, there remain unresolved problems. Nonetheless, even with controversial and contradictory evidence casting some doubt, it is relatively safe to state that some points have been established:

*Developing organisms.* The following data seem to be indicative of some of the radiation effects on developing organisms.

(1) Deficiencies in maze learning occur in rats irradiated on any gestation day and during the neonatal period. These deficiencies can be elicited by a dose as low as 25r. The extent of the deficit is a function of the time and amount of radiation given and the age at time of testing, the type of maze employed, and the sex of the animal. In general, maze-learning deficiency is more apparent for sublethal doses (100-200r) delivered *in utero* around the fourteenth day. Testing of animals for acquisition of maze-learning skill when they are around one year of age reveals greater deficit than appears in rats tested at an earlier age. The female is more adversely affected than the male; a Lash-



ley III type maze, which has been shown to be especially sensitive for studying brain damage (Lashley 1929), is the instrument used to obtain the findings.

(2) Alterations in activity and locomotion result from prenatal and neonatal exposure with doses as low as 25r. Radiation *in utero* seems to enhance these behaviors, while neonatal exposure inhibits them.

(3) Prenatally irradiated subjects exhibit greater "fearfulness," a measure of emotionality, than do normals. A 50r dose seems to be sufficient to elicit this behavior.

(4) Although there are reports of a few cases of personality aberrations and mental deficiencies in Japanese irradiated *in utero*, as a result of the bombing of Hiroshima and Nagasaki, there is insufficient data available to draw conclusions regarding this population (Miller 1956).

*Adult organisms.* For adult organisms some of the effects of radiation are:

(1) It has been generally accepted that immediately following doses in sublethal and even lethal ranges, higher-order learning processes and their retention are not impaired (Northwestern Univ. 1962; Furchtgott 1963). For example, Kaplan and his associates (see Pickering, Langham, and Rambach 1956) have shown that after exposure to massive doses of gamma radiation, monkeys were able to learn to discriminate between stimuli to which they had never before been exposed. These same animals manifested memory for discriminations they had learned to make before they were irradiated. Exceptions have been reported (Sharp & Keller 1965) which indicate that monkeys exposed to doses which cause death within hours will show impairment of preradiation-learned performances which involve memory of time perception. If the irradiated subject survives beyond a thirty-day period the dose is then not viewed as lethal. For animals receiving nonlethal doses, there is evidence after months and years of changes in performance which are attributable to radiation exposure (Davis 1965). Among the changes noted are alterations in motivation, extent of distractibility, and narrowing of attention, all factors related to higher-order learning and retention.

(2) Drive is adversely affected by radiation exposure, in direct association with the reduction of food and water intake and the decrement in general activity.

(3) Radiation has been shown to be effective as an unconditioned stimulus. Conditioned avoidance responses can be elicited with as little as 10r, providing a means for testing conditioned responses to sound, light, taste, and object quality.

(4) Sublethal doses up to 1,100r have resulted in increases in self-involved, and a diminution in socially oriented, behavior of survivors. This finding was consistent during a period of seven years of observation of subjects.

(5) Low doses of X rays have been shown to be visible. Also, it appears that visual functions are depressed by radiation.

(6) Cutaneous, gustatory, olfactory, and interoceptive mechanisms are found to have reversible and unstable alterations of response to radiation. Auditory responses, also, reflect these characteristics, with conflicting reports of both increases and decreases of threshold.

(7) Very low doses have been shown by some Russians to alter conditioned reflex behavior (Stahl 1959; 1960; Lebedinskii et al. 1958).

*Neurophysiological aspects.* Since radiation psychology falls quite logically under the broader fields of physiological psychology and radiobiology, attention should be directed to interrelationships between neurophysiological processes affected by radiation and related behavioral adjustments. The body of biological data is extensive and may be studied in the previously mentioned source books. Some of the most relevant findings can be seen to relate closely to behavioral results:

(1) The relative radiosensitivity and resistance of various cells and tissues have been identified, with strong evidence that the greatest sensitivity is in the lymphocytes and granulocytes and the least sensitivity is in the muscle, bone, and nerve cells.

(2) Radiation impairs the ability of the cells to maintain metabolic equilibrium and ultimately provides evidence for cell deviancy both morphologically and functionally.

(3) For developing organisms, the anatomical changes are maximal if radiation occurs during the period of major organogenesis. There is also considerable morphological data to support the conclusion that radiation delivered at any time to a developing organism has a deleterious effect upon neurophysiological function.

(4) It is particularly relevant to mention the many neurological signs observed in organisms after neonatal radiation: tremor, leg dragging, head deviations, uncoordinated gait, elevated pelvic posture, and chronically abducted limbs.

(5) Adult neural tissue has for the most part been established as being markedly radioresistant, with doses in the 1,000r level failing to create impairment. Nonetheless, a growing body of data from European and Russian laboratories gives functional evidence of neural impairment after minute doses.

(6) Innumerable references are available that

reflect radiation effects upon the interaction of the nervous system with endocrine, cardiovascular, and gastrointestinal functions. All of these factors are strongly related to the behavioral findings noted.

It is too early in the development of the field of radiation to predict the direction it will take. However, evidence does indicate that radiation can be a useful tool in the study of embryonic development (Hicks 1958). Suggestions are being made for the exploration of the use of radiation as a diagnostic tool in clinical psychology and psychiatry, through employment of radioisotopes in chemotherapy and physical therapy. Its usage in studying electronically developed brain models points to a new behavioral domain of investigation allied to computer technology. The need for exploring the sociological implications of a radiated society demands development of behavioral science techniques for identifying radiation factors that interact with other group attributes.

Perhaps the most important need in the field today is to standardize procedures when findings appear to be consistent. The reduction of controversial data to a minimum will permit generalization within psychology and between the other disciplines in this field. The fact that so many disciplines are linked to the radiation field demands this type of liaison and interdisciplinary cooperation if the massive task of integrating the knowledge of radiation already obtained in the atomic era is to be accomplished.

SYLVAN J. KAPLAN

[Other relevant material may be found in NERVOUS SYSTEM; NUCLEAR WAR; SPACE, OUTER.]

#### BIBLIOGRAPHY

- BOND, VICTOR P.; FLIEDNER, THEODORE M.; and ARCHAMBEAU, J. O. 1965 *Mammalian Radiation Lethality: A Disturbance in Cellular Kinetics*. New York: Academic Press.
- DAVIS, ROGER T. 1965 The Radiation Syndrome. Volume 2, pages 495-524 in Allan M. Schrier, Harry F. Harlow, and Fred Stollnitz (editors), *Behavior of Non-human Primates: Modern Research Trends*. New York: Academic Press.
- ELLINGER, FRIEDRICH P. (1935) 1947 *The Biological Fundamentals of Radiation Therapy*. London: Cleaver-Hume. → First published in German.
- FURCHTGOTT, ERNEST 1956 Behavioral Effects of Ionizing Radiation. *Psychological Bulletin* 53:321-334.
- FURCHTGOTT, ERNEST 1963 Behavioral Effects of Ionizing Radiations: 1955-1961. *Psychological Bulletin* 60: 157-199.
- FURCHTGOTT, ERNEST 1964 Postscript to "Behavioral Effects of Ionizing Radiations: 1955-1961." *Psychological Bulletin* 61:236-238.
- GLASSER, OTTO et al. (1944) 1961 *Physical Foundations of Radiology*. 3d ed., rev. & enl. New York: Hoeber.
- GRAY, L. H. 1949 Comparative Studies of the Biological Effects of X-rays, Neutrons, and Other Ionizing Radiations. Pages 114-137 in N. Howard-Jones (editor), *Applied Biophysics; Survey of Physical Methods Used in Medicine: A Symposium*. New York: Chemical Publishing Company.
- HANRAHAN, JAMES S.; and BUSHNELL, DAVID 1960 *Space Biology: The Human Factors in Space Flight*. New York: Basic Books.
- HICKS, S. P. 1958 Radiation as an Experimental Tool in Mammalian Developmental Neurology. *Physiological Reviews* 38:337-356.
- HOLLAENDER, ALEXANDER (editor) 1954-1956 *Radiation Biology*. 3 vols. New York: McGraw-Hill. → Volume 1: *High Energy Radiation*. Volume 2: *Ultraviolet and Related Radiations*. Volume 3: *Visible and Near-visible Light*.
- INTERNATIONAL SYMPOSIUM ON THE RESPONSE OF THE NERVOUS SYSTEM TO IONIZING RADIATION, SECOND. LOS ANGELES, 1963 1964 *Response of the Nervous System to Ionizing Radiation*. Edited by Thomas J. Haley and Ray S. Snider. Boston: Little.
- KAPLAN, SYLVAN J. 1962 Radiation Research in Psychology: An Analysis of Techniques in Maze Experimentation. *Psychological Bulletin* 99:153-160.
- KIMELDORF, DONALD J.; and HUNT, EDWARD L. 1965 *Ionizing Radiation: Neural Function and Behavior*. New York: Academic Press.
- LAPP, RALPH E.; and ANDREWS, HOWARD L. (1948) 1963 *Nuclear Radiation Physics*. 3d ed. Englewood Cliffs, N.J.: Prentice-Hall.
- LASHLEY, KARL S. 1929 *Brain Mechanisms and Intelligence: A Quantitative Study of Injuries to the Brain*. Univ. of Chicago Press.
- LEA, DOUGLAS E. (1947) 1955 *Actions of Radiations on Living Cells*. 2d ed. Cambridge Univ. Press.
- LEBEDINSKII, ANDREI V.; and NAKHIL'NITSKAIA, Z. N. 1963 *Effects of Ionizing Radiation on the Nervous System*. Amsterdam and New York: Elsevier.
- LEBEDINSKII, ANDREI V. et al. 1958 On the Biological Effect of Small Doses of Ionizing Radiation. Pages 17-28 in International Conference on the Peaceful Uses of Atomic Energy, Second, Geneva, *Proceedings*. Volume 22: Biological Effects of Radiation. Geneva: United Nations.
- MILLER, R. W. 1956 Delayed Effects Occurring Within the First Decade After Exposure of Young Individuals to the Hiroshima Atomic Bomb. *Pediatrics* 18:1-18.
- NORTHWESTERN UNIVERSITY, EVANSTON, ILLINOIS; and MEDICAL SCHOOL, CHICAGO 1962 *Response of the Nervous System to Ionizing Radiation: Proceedings of an International Symposium*. New York: Academic Press.
- PICKERING, JOHN E.; LANGHAM, WRIGHT H.; and RAMBACH, WALTER A. 1958 *The Effects From Massive Doses of High Dose Rate Gamma Radiation on Monkeys* USAF, Aerospace Medical Center (ATC), Report No 60 57. Brooks Air Force Base, Texas: School of Aviation Medicine.
- RUGH, R. 1962 Ionizing Radiations and Congenital Anomalies of the Nervous System. *Military Medicine* 127:883-907.
- RUGH, R. 1963 The Impact of Ionizing Radiations on the Embryo and Fetus. *American Journal of Roentgenology, Radium Therapy and Nuclear Medicine* 89 182-190.
- RUSSELL, LIANE B. 1954 *The Effects of Radiation on Mammalian Prenatal Development*. Volume 1, pages 861-918 in Alexander Hollaender (editor), *Radiation Biology*. New York: McGraw-Hill.



- SHARP, JOSEPH C.; and KELLER, BARBARA K. 1965 *A Comparison Between the Effects of Exposure to a Mixed Fission Spectrum Delivered in a Single "Pulse" and X-rays Delivered at a Slower Rate Upon the Conditioned Avoidance Behavior of Primates*. Technical Report No. 4. Washington: Walter Reed Army Institute of Research, Division of Neuropsychiatry.
- STAHL, WALTER R. 1959 A Review of Soviet Research on the Central Nervous System Effects of Ionizing Radiations. *Journal of Nervous and Mental Disease* 129: 511-529.
- STAHL, WALTER R. 1960 Recent Soviet Work on Reactions of the Central Nervous System to Ionizing Radiation. *Journal of Nervous and Mental Disease* 131: 213-233.
- VAN CLEAVE, CHARLES D. 1963 *Irradiation and the Nervous System*. New York: Rowman & Littlefield.
- WORLD HEALTH ORGANIZATION 1957 *The Effect of Radiation on Human Heredity*. Geneva: The Organization.

## RADICALISM

The term "radicalism" always points to some analytical or revisionist function. It implies a concentration of the focus of relevance on a particular principle, at the expense of the traditionally sanctioned regard for the complexities of context. The element thus abstracted becomes the salient core on which inference and action are based. Radicalism tends to be comprehensive; no matter where it starts, it tends to assimilate all aspects of life to the initial principle. In its positive sense this tendency implies a projection of a completely new version of human life and enterprise. In its negative sense it implies a threat to all aspects of ongoing life. Although many forms of radicalism eschew violence, there is little doubt that the overthrow of the existing order is part of the radical agenda. The radical reduction and its extension into a comprehensive doctrine need not have the character of logically derived conclusions. Quite often the assertions are the result of directly intuited truth, and the forms of expression range from scholastic demonstrations to rhapsodic prophecy. Whatever the form, the impulse behind it is to announce the sovereignty of a principle and to render a principled, unified, and internally consistent interpretation of the cosmos and the meaning of human life. It is expected that the believer's perception of and attitude toward reality will be synthesized accordingly into an internally coherent outlook and ultimately translated into a principled conduct of life.

**Zealotry and rationality.** There exists a close but somewhat misleading affinity between radicalism and zealotry, particularly in the areas of politics, morality, and religion. Zealotry, being a form of unquestioned and permanent obeisance, may, and

often does, become the routinized pursuit of radical ideals. The zealot's commitment to radical causes is accompanied by deep-seated emotional problems. The particular psychological structure of motives and attitudes that leads to or is compatible with expressions of zealous radicalism has been identified as the "authoritarian personality type" (Adorno et al. 1950).

However, even when the attitude of the radical is wholly immersed in feeling, its expression is not compatible with emotional impulsiveness but is always pervaded by its own peculiar rational discipline. The rational structure of radical thought and action is the expressive mold that fits the unconscious-need dispositions of many persons; to others it is a matter of deliberate moral choice and closely reasoned inference; and to some it may be no more than a mask they can plausibly and deliberately display while being motivated by considerations of momentary expediency.

The conception of radicalism as rational does not imply that it is ruled by logic, science, or economy but merely that the drive toward some sort of explicit intellectual generalization of the meaning of human action and experience is a constitutive property of radicalism. By methodically correlating its interpretive principles with its maxims of conduct, radicalism provides ultimate, permanent, and "objectively valid" grounds for moral choice. Radicalism usually attaches its single-minded emphasis to something that is already known and valued, albeit as one thing among many. What is new is not the tenet itself but the militant assertion of its sovereign and unqualified supremacy. However, while the banner of radicalism derives from the outlook of the society in which it occurs, its arguments turn against that society.

Although radicalism is characteristically rational or at least rationalized, it is clearly and implacably inconsistent with reasonableness. In principle, pure radical thought and action is entirely devoid of practical wisdom, of sensitivity for the occasion, of opportunistic economizing, of the capacity to learn from experience, of flexibility and looseness of interest. In sum, it lacks that bargaining side of intelligence that characterizes the conduct and thinking of "reasonable" persons. Indeed, the principal polemic adversary of the radical is the normally competent, fully franchised member of his society whose orientation to reality is governed by conventional standards of "reasonableness." While the latter moves in a world of ambiguity and uncertainty with apparent ease and comfort, varying the perspective of his interest and judgment in accordance with practical considerations and chang-

ing circumstances, the radical finds in such behavior decadence and philistinism. Because of the radical's dedication to the salience of his reduction, he is, qua radical, divested of realistic concern for the passing complexities of everyday life, and he views them with contempt. He scorns the demands that issue from "the way things are," and eschews partaking in the respect that is accorded to persons and arrangements on purely conventional grounds.

**The radical ideology.** It has been shown that radicalism has its origins and permanent focus of appeal in the socially displaced strata of society (Lipset 1960). Although this phenomenon is probably not present in every form or instance of radicalism, it is immensely important for the understanding of the social structure of radicalism. The socially eccentric locus of origin of radicalism appears to be a particularly auspicious perspective for the deliberate appraisal of the ordinarily taken-for-granted texture of everyday life, and it may well be that man's consciousness of himself and of the world around him is the assimilated aggregate of past and forgotten radicalizations.

When men of radical persuasion look into the past, one of two alternative views is obtained. In one, the tendencies contained in mundane history are disregarded entirely in favor of tracing some doctrinal claim to a distant and mythical past. All that happened between some golden age and the radical discernment in its modern setting is defined as an ironic departure from the intended direction of the original state. In the second and more common view, mundane history is seen as a script of a determined sequence of events. In the huge historical panoramas of Hegel and Marx, all that is truly historical is defined as accidental and insignificant, and all that matters is the permanently lawful sequence of general tendencies. What Hegel called *die Tücke des Geistes* merely deludes man into thinking that what he does as a free agent matters historically. From either point of view, the polemic highlight of radical historiography is the announcement of an eternal truth. In either case, the image of the past urges that the full meaning of man's misery and degradation cannot be fully grasped by beholding it now and that the proposed remedies cannot be fully justified without considering them as links in a process of cosmic proportion. Knowledge of the past has as its principal function the exemplification and the aggrandizement of the timeless truth contained in the radical doctrine and the demonstration of its compelling necessity.

The doctrine and the ethic of radicalism is inevitably invested with tensions. On the one hand, it is systematized by internal standards of warrant

and sensibility; on the other hand, radical beliefs remain responsive to tests of everyday life experiences. Even while the great goals and the order of the future are within the grasp of the imagination, there remain the petty vexations and distractions of everyday life. Even while he holds steadfast to the grand scheme, the believer must resist temptations to suspend the relevance of the doctrine for considerations of momentary interest. Worse yet, his resoluteness is threatened continuously by counterevidence that tends to discredit the doctrine. In the long run radicals succumb to these worldly pressures; they pay their price in sacrifice, settle for partial gains, and become assimilated. Often when they surrender their totalitarian claims, they become selectively dogmatized and develop the art of casuistry. This process is best exemplified in the transformation of sects into churches. The alternative to the corruption of rational purity and consistency, i.e., the maintenance of radicalism in its pristine state, requires that radical movements impose upon their members a form of discipline that makes doctrinal impurity as difficult and unattractive as possible and that peremptorily discredits the relevance of all possible distraction and counterevidence.

Several well-recognized features of extreme radical movements have precisely this effect. Most noteworthy are the following: (1) The believers are organized into a charismatic fellowship. Within this collectivity the tenets of the faith can be exercised freely and, thus, are validated, i.e., the fellowship is "a living example." (2) the distinctness of the radical creed and program from the rest of the world is symbolically emphasized. A common device to accomplish this is to associate it with the inspiration of a prophet. The particularistic access to truth discredits a variety of polemic opponents. (3) There is a sustained concern for the purity of belief and conduct. The state of permanent purge functions at the level of self-critique, as well as that of collective and authoritative discipline. (4) The manner in which the doctrine pertains to the lives of believers is diffusely comprehensive. All interest and activity is "normalized" through participation in the movement. Not only does the movement usurp enormous powers of restraint over its members; it also annuls the significance of all external sources of sanction. (5) The movement monopolizes the commitments of the believers. Loyalty to the movement requires that no personal interest or obligation may be admitted as legitimately contesting a demand issuing from participation in the movement. This condition is substantially satisfied by dissolving all possible human ties



and by deindividualizing members in the direction of some heroic ideal. (6) Suffering and martyrdom must be made acceptable and be brought within the purview of immediate possibility. While members are desensitized to pain, their conception of the relationship between the assailant and the victim becomes impersonal, and brutality becomes morally neutral. (7) The movement exploits outside sentiments against it to its own organizational advantage. That is, members who are publicly compromised by participation in it are forced to burn their bridges behind them.

### The development of modern radicalism

The original cause with which modern radicalism is associated is the attack on the traditionally inherited corporate structure of power, in the name of an equal and liberal distribution of political franchise—i.e., the ideals of democracy. Although movements of democratic reform and rebellion appear sporadically throughout history, radicalism in this sense became endemic only to the Western world, and only since the late eighteenth century. During the Reformation, the main Protestant denominations, notably the Evangelical Lutheran church and the Church of England, never lost their footing in the established political hierarchy and have settled their quarrels with temporal authority under the doctrine, *Cuius regio eius religio*. Other sectarian movements, however, have been sporadically involved in partisan warfare against traditional secular privilege. The opening and the closing of this period of complicated plebeian upheavals are marked by the *Bauernkriege* in sixteenth-century Germany and the Levellers' movement of the Puritan revolution in England. In 1685, at the time of the suspension of the Edict of Nantes, the fight for religious and political freedom was largely exhausted, muzzled, or confined to insignificant enclaves, from which irreconcilable partisans embarked on their voyages to the New World. Even in eighteenth-century England the dissenting bodies in the Bunyan tradition became reconciled to the *status quo*.

Founded in part upon the largely academic discourses of Locke and Montesquieu, the doctrine of modern democratic radicalism was formulated, by Rousseau and a host of lesser writers, with full awareness of its seditious character. Their relentless argument urged that all existing conditions and customs are unnatural and must be destroyed and replaced by a new and rational order (Talmon 1952). In contrast to the religiously inspired radicalism of the Reformation period, the new formulation was wholly secular.

Foremost among the objects of attack was the belief, never seriously challenged in previous ages of Western civilization, that only a select few, an elite, had the wisdom and right and power to govern. In its place there was formulated a doctrine that proclaimed every man the patron of his own life and established the sovereign right of all the people to order their common affairs. The vision of individual liberty and popular sovereignty was made compelling by the endorsement of reason. Although it is now easily seen that this enlightenment of reason contained a strong admixture of romantic sentimentality, there is little doubt that the protagonists themselves perceived their arguments as appealing to discretion rather than feeling. Of course, the radicals did not invent or monopolize the use of reason as an instrument of politics. Other forms of governing and the opposition to radical innovation also had their rational apologists. None, however, depended to the extent the democratic radicals did on the persuasive power of reason and excluded so completely other considerations.

**Early radical movements.** The American and French revolutions in the closing decades of the eighteenth century were significant expressions of modern radicalism. They did not, however, express all existing radical tendencies, and they encompassed a great deal more than merely radical idealism. More specifically, five movements, all derived from ideals of the enlightenment, show the directions and development of modern radicalism.

**Jacobinism.** The most direct translation of enlightenment ideals into political action was Jacobinism, which had been effectively practiced primarily in France—although Jacobin clubs existed in most European countries and in the United States—and for only a relatively short period of time. Yet it made a permanent contribution to the subsequent organization of all radical agitation. The small, locally based Jacobin clubs were focal points for revolutionary propaganda campaigns and were employed to mobilize mass support whenever the small core of activists needed such support. Furthermore, the tight network of relations within the clubs and the efficient network of communications between clubs provided for a state of permanent purge that made Robespierre's slogan about the tyranny of liberty a practical possibility. Jacobinism is an early model of a movement that requires for its perpetration a very high level of activist tension.

**Populism.** Populism assumed a variety of forms, depending on the place of its occurrence.

In the United States it traces back to the ideals of Jeffersonian democracy. Despite Jefferson's own cosmopolitan mind, these ideals contained physiocratic elements that in the hands of Jackson became a permanent radical trend in American politics. The core of strength of American Populism was always agrarian, but Populism directed some of its appeals to, and derived some of its support from, the urban masses. In Russia a variety of influences, most notably Slavophile sentiments, led to the formation of the Narodnik movement around the time of the emancipation of the serfs. Here imported enlightenment ideals combined with romantic nationalism in the revival of the concept of the *mir*, an ancient Slavic form of agricultural polity. In Germany, starting with Herder, scholars and writers with a romantic medievalist bent cultivated an interest in folklore, ethnic history, and ultimately, national sovereignty. Although the German development concentrated on the intellectual glorification of the concept of *Volk* and did not lead to the formation of a populist movement in Germany, it furnished the ideological underpinning for a number of populist-peasant parties in eastern and central Europe. The common element of all these trends was a strong belief in the rights and creative powers of the common man, living close to nature, whose interests are naturally opposed to the oligarchic tendencies of large central governments and to professional political administration.

*Philosophical radicalism.* The ideals of Jeffersonian democracy and those of the philosophical radicals of England are similar in that historically both are of Whig origin. The English reformers, however, were of distinctly bourgeois persuasion and appeal. Their idea of reconstructing government according to the principle of utility was an exercise in business rationality. Although the ideals of Bentham, the leading theoretician of this group, were close to the ideals of the Jacobins, and although he was in his own way as much a rationalistic simplifier as they were, the characteristic difference between them is that for the English reformer the fight for a rational social order did not admit the possibility of violence and chaos, even as a tactical instrument [see BENTHAM].

*Anarchism.* The strongest and most systematically radical version of radicalism is to be found in anarchism. William Godwin first formulated the complete version of anarchism, teaching that to compel men to act according to reason is superfluous and to compel them to act against reason is unjust. On this basis, he called for the abolishment of all institutions. With Proudhon, anarchism

adopted a program of economic reform that far exceeded anything contained in other contemporary radical tendencies. Later the movement assumed the character of a quasi political party, always, however, refusing to participate in the affairs of government [see ANARCHISM].

*German idealism.* German idealism has been called "the theory of the French Revolution." Starting with the powerful influence of Rousseau on Kant and continuing through Schelling and Fichte, this tendency culminated in the work of Hegel. Here the ideas of freedom and reason were wholly emancipated from man as a concrete being and were objectified in the modern state. Hegel's statism is the perfect obverse of anarchism. Whereas for Godwin man in his natural state is the paradigm of reason and freedom, for Hegel reason and freedom are the attributes of a transcendental subjectivity whose only possible concrete manifestations are institutions. Although Hegel's own thinking turned to the glorification of the existing Prussian state, the logic of his argument exerted a powerful influence on the formation of all subsequent radical doctrines.

From its beginning and through the first few decades of the nineteenth century, modern radicalism had a certain enthusiastic unity, expressed in the slogan, *Pas d'ennemis à gauche!* Even Fichte and the young Hegel stood on the left, in opposition to the conservative restoration forces of the post-Napoleonic era. From approximately the middle of the nineteenth century on, there appeared a split into what came to be known as right-wing radicalism and left-wing radicalism. The split was first manifest among the disciples of Hegel and ultimately led to the realignment of all radical forces.

*Right-wing radicalism.* By and large, right-wing radicalism is not readily discernible in nineteenth-century politics. The consolidation of the European states followed closely the Hegelian blueprint, and right-wing radicalism appeared mainly in the form of programs seeking to perfect an already existing state of affairs. Revisionist radical right-wing tendencies appear only at the close of the century, through the infusion of a new element. Statism achieved a new formulation, in a doctrine that based national sovereignty on historical missions of distinct races. The principal protagonists of this idea were Joseph Arthur de Gobineau and Houston Stewart Chamberlain. Although this movement explicitly repudiated the influence of Hegel, it drew heavily on the translation of the idea of *Weltbürgerthum* into the idea of the *Nationalstaat*, and there is little doubt that it absorbed the bulk of the



forces that were earlier affiliated with the Hegelian right [see GOBINEAU].

The association between ethnic identity and territorial sovereignty became universally accepted during the nineteenth century and furnished the basis for the political division of Europe at Versailles. During the two decades following World War I, nationalism achieved its most radical expression in the totalitarian dictatorships of Italy, Japan, and Germany. The doctrinal underpinning of these systems varied, emphasizing alternatively race, cultural heritage, or a combination of both. Associated with them are attitudes of superpatritism and a belief in the superiority of one's ethnic group, and they encompassed policies of belligerent imperialism. Movements of this sort were active and continue to be active in virtually all nations of the world, and many autocrats have found in the rhetoric of radical nationalism the justification for claiming power [see FASCISM; NATIONAL SOCIALISM].

A number of phenomena of somewhat lesser significance deserve mention in the context of right-wing radicalism. In France the Radical Republicans stood in opposition to the "opportunistic" left wing of the Parliament of 1875 and subsequently became allied with militarist, proclerical, and royalist groups. This alliance was consolidated during the Dreyfuss affair and remained a sporadically active extreme right-wing faction in French politics. Phenomena such as the Falange in Spain and the Black Guard in Rumania also belong to this category [see FALANGISM]. Standing apart from the above but related to them are anti-civil liberties tendencies in American politics. The latter include, among others, such heterogeneous elements as the radicalism of Thaddeus Stevens and his followers, certain nativist movements, and the outlook referred to as McCarthyism.

**Left-wing radicalism.** The structural affiliation of right-wing radicalism with established political power led to its development as the philosophy of the existing polity, i.e., as an apologia. Hence, the term "radicalism" came to mean left-wing radicalism. But from its very onset this movement was torn by schisms. The last left-wing synthesis, during the two decades following the revolution of 1848, encompassed some socialists of the Jacobin heritage, some English reformers of the labor-class movement, the anarchists under the leadership of Bakunin, disciples of Moses Hess and Ferdinand Lassalle, and, of course, Marxists. By the time of the Paris Commune, this shaky alliance was in shambles. Since then, left-wing radicalism has been divided into numerous factions, which have often fought each other with bitter hostility.

Anarchists have maintained their intransigent alienation from normal political processes. Occasionally they have displayed some strength in the organization of labor movements (e.g., the Industrial Workers of the World in the United States) or in Latin-European politics (e.g., in Spain during the 1930s). On the whole, however, anarchism has remained conspicuous mainly because of real or alleged acts of terrorism. Its occasional associations with criminal elements, i.e., forces inimical not only to the established order but to anarchist ideals themselves, have caused anarchism to be generally regarded as the most virulently destructive form of radicalism.

Marxist communism achieved a strong international party organization, maintaining an orthodox revolutionary line, and gained political ascendancy in several European states in the aftermath of World War I. Of these, the Soviet Union is the only one that remained stable. Following World War II a number of states within the traditional sphere of political influence of Russia came under the domination of the Communist party. These governments instituted nationwide collectivization of productive property and strong central organizations of power, in strict opposition to anarchist teachings.

Evolutionary socialism, which is in part a continuation of English philosophical radicalism and in part an offshoot of Marxism, worked toward the assimilation of elements of the radical doctrine into the mainstream of twentieth-century politics, thereby eliminating the very conditions that were the foundation for the Marxist prophecy. Some socialist parties include the word "radical" in their party designation with justification. Others, like the French Radical Socialists under the leadership of Paul Herriot, are more properly located in the center of the general political spectrum.

Of the many lesser schismatic left-wing radical movements, syndicalism is perhaps the most important. Although mainly a hybrid offshoot of Marxist and anarchist influences, it contained some original ideas and retained a distinct identity over a protracted period of time. Syndicalism exercised a strong influence in some Latin countries of Europe and the Americas, and contributed, *inter alia*, to the formulation of the doctrine of Italian fascism [see SYNDICALISM].

Radicalism is still very much alive everywhere, but particularly in those countries that recently became free of European domination. It is in the nature of the situation that those who are in this fight must share the forum and spotlight with others, who irresponsibly exploit vulgar frustra-

tions. But even as the fight against the old tyrannies continues, the very polity that radicalism has built has become the focus of new discontents. The conscious and deliberate effort to bring the human community under the rule of reason has led to the development of complex institutions. Intended to insure and increase freedom, equality, and popular welfare, these institutions have recently been attacked as a new, dehumanized tyranny. The theme of alienation is heard again after a century of dormancy, and on the fringes of society there are groups that view the existing social order as their antagonist. Thus, radicalism, which started by abolishing all sanctions of history, has returned to the fold of history, partly by its triumph and partly by its own sins. While the incorporation of the ethics of political freedom and universal rational justice into human consciousness appears to be a permanent achievement, the shadow of new radical questions about the human condition is cast upon the stage of human life.

Radicalism is a part of the general theme of the growth of rationalistic ethics. To the extent that Western civilization has been the most fertile soil for this development, radicalism is native to the West. Of course, neither rationalism nor radicalism is an Occidental monopoly; nor is Western history the record of a sustained development of rational ethics. However, as Max Weber pointed out, the prophetic origin of the Judaeo-Christian tradition, the development of a rational economy, the growth of rational science and philosophy, the organization of authority along legal-rational lines, together testify to the fact that reason is the guiding demon of Western man. But the search for an intellectual synthesis of the meaning of life forever transcends life. Reason, like myth, in attempting to grasp and express the roots of being, always idealizes the realities of everyday existence. In radicalism the ideal is the supreme taskmaster. The tension between the activist ethical principle and the exigencies of existence is the core of the radical dilemma.

EGON BITTNER

[See also ANARCHISM; CONSERVATISM; MARXISM; NATIONALISM; PERSONALITY, POLITICAL; SOCIAL MOVEMENTS; SOCIALISM; TOTALITARIANISM. Also related are the biographies of BAKUNIN; FANON; MARX.]

#### BIBLIOGRAPHY

- ADORNO, THEODOR W. et al. 1950 *The Authoritarian Personality*. American Jewish Committee, Social Studies Series, No. 3. New York: Harper.
- AMERICAN ACADEMY OF ARTS AND SCIENCES 1954 *Totalitarianism*. Edited with an introduction by Carl J. Friedrich. Cambridge, Mass.: Harvard Univ. Press. → Proceedings of a conference held at the American Academy of Arts and Sciences, March 1953.
- BELL, DANIEL (editor) (1955) 1964 *The Radical Right*. Garden City, N.Y.: Doubleday. → An expanded and updated edition of *The New American Right*.
- BITTNER, EGON 1963 *Radicalism and the Organization of Radical Movements*. *American Sociological Review* 28:928-940.
- BRINTON, CLARENCE CRANE (1930) 1961 *The Jacobins: An Essay in the New History*. New York: Russell.
- COHN, NORMAN (1957) 1961 *The Pursuit of the Millennium: Revolutionary Messianism in Medieval and Reformation Europe and Its Bearing on Modern Totalitarian Movements*. 2d ed. New York: Harper.
- FISH, CARL R. (1927) 1946 *The Rise of the Common Man: 1830-1950*. New York: Macmillan.
- GEYER, CURT 1923 *Der Radikalismus in der deutschen Arbeiter Bewegung: Ein soziologischer Versuch*. Jena (Germany): Thüringer Verlagsanstalt und Druckerei.
- HALÉVY, ÉLIE 1901-1904 *La formation du radicalisme philosophique*. 3 vols. Paris: Alcan.
- HOFFER, ERIC 1951 *The True Believer: Thoughts on the Nature of Mass Movements*. New York: Harper. → A paperback edition was published in 1958 by New American Library.
- KORNHAUSER, WILLIAM 1953 *Liberal and Radical Political Careers: A Study of Group Loyalty*. Ph.D. dissertation, Univ. of Chicago.
- LANDAUER, CARL 1959 *European Socialism: A History of Ideas and Movements From the Industrial Revolution to Hitler's Seizure of Power*. 2 vols. Berkeley: Univ. of California Press. → Volume 1: *From the Industrial Revolution to the First World War and Its Aftermath*. Volume 2: *The Socialist Struggle Against Capitalism and Totalitarianism*.
- LIPSET, SEYMOUR M. 1960 *Political Man: The Social Bases of Politics*. Garden City, N.Y.: Doubleday.
- MANNHEIM, KARL (1929-1931) 1954 *Ideology and Utopia: An Introduction to the Sociology of Knowledge*. New York: Harcourt; London: Routledge. → First published in German. A paperback edition was published in 1955 by Harcourt.
- MARCUSE, HERBERT (1941) 1955 *Reason and Revolution: Hegel and the Rise of Social Theory*. 2d ed. New York: Humanities; London: Routledge. → A paperback edition was published in 1960 by Beacon.
- MEINECKE, FRIEDRICH (1924) 1962 *Machiavellism: The Doctrine of Raison d'État and Its Place in Modern History*. New York: Praeger. → First published as *Die Idee der Staatsräson in der neueren Geschichte*. Contains a general introduction to Friedrich Meinecke's work by W. Stark.
- MICHEL, ROBERT 1925 *Psychologie der antikapitalistischen Massen Bewegungen*. Section 9, part 1, pages 241-359 in *Grundriss der Sozialökonomik*. Tübingen (Germany): Mohr.
- MIŁOŚĆ, CZESŁAW (1953) 1955 *Captive Mind*. New York: Vintage. → First published in Polish.
- NICOLET, CLAUDE 1957 *Le radicalisme*. Paris: Presses Universitaires de France.
- PALMER, ROBERT R. 1959-1964 *The Age of the Democratic Revolution: A Political History of Europe and America, 1760-1800*. Princeton Univ. Press.
- PÖHLMANN, ROBERT VON (1893-1901) 1925 *Geschichte der sozialen Frage und des Sozialismus in der antiken Welt*. 3d ed. Munich: Beck. → First published as *Geschichte des antiken Kommunismus und Sozialismus*.
- POLANYI, MICHAEL 1960 *Beyond Nihilism*. Cambridge Univ. Press.



- SELZNICK, PHILIP (1952) 1960 *The Organizational Weapon: A Study of Bolshevik Strategy and Tactics*. Glencoe, Ill.: Free Press.
- SHILS, EDWARD 1954 Authoritarianism: "Right" and "Left." Pages 24-49 in Richard Christie and Marie Jahoda (editors), *Studies in the Scope and Method of The Authoritarian Personality*. Glencoe, Ill.: Free Press.
- SHILS, EDWARD 1956 *The Torment of Secrecy: The Background and Consequences of American Security Policies*. Glencoe, Ill.: Free Press.
- SOMBART, WERNER (1896) 1909 *Socialism and the Social Movement*. London: Dent; New York: Dutton. → First published as *Sozialismus und soziale Bewegung im 19. Jahrhundert*. Title later changed to *Sozialismus und soziale Bewegung*. The 10th edition was published in 1924 as *Der proletarische Sozialismus ("Marxismus")*.
- TALMON, JACOB L. (1952) 1965 *The Origins of Totalitarian Democracy*. 2d ed. New York: Praeger.
- VENTURI, FRANCO (1952) 1960 *Roots of Revolution: A History of the Populist and Socialist Movements in Nineteenth Century Russia*. New York: Knopf. → First published as *Il populismo russo*.
- WILSON, EDMUND (1940) 1953 *To the Finland Station: A Study in the Writing and Acting of History*. Garden City, N.Y.: Doubleday.

### RADIN, PAUL

Paul Radin, American anthropologist, was born in Lodz, Russian Poland, in 1883 and died in New York City in 1959. Despite the fact that he was brought to the United States in infancy, he never severed his European roots, and he could never bring himself to accept the New World completely, except, ironically, for his association of nearly fifty years with the Winnebago Indians of Wisconsin.

Radin's personal and social background was complex. He came from a German-Russian Jewish family that had become secularized in the mode of the *haskalah*, the Jewish strand in the western European Enlightenment. His father was a rabbi of the reform movement, a Hebrew scholar, and a linguist. Herman, his oldest brother, became a physician, and Max, next in line, was a distinguished legal scholar. For the Radins these secular professions, which refocused the traditional Jewish concern with ritual learning but which did not abandon learning as a ritual, were ideologically correlated with a humane, if skeptical, liberalism. The skepticism and rationalism, however, were largely reactive protests against the cramping social and intellectual orthodoxy of the patriarchal and theocratic past. Yet the passion for scholarship, the commitment to human realization in the world, and, in Radin's case, the fascinated concern with religion and ethics, maintained a distinc-

tively Jewish cast. Moreover, Radin's intellectual cosmopolitanism, his radicalism, his conception of learning as a moral enterprise, and his capacity to live almost exclusively the life of the mind, making him, therefore, dependent upon friends for a variety of services, represent further elements that characterized the Jewish scholar *en passage* from the traditional milieu to the modern industrial and urban world. But Radin never came to terms with modern technology; he never learned to use a typewriter, for example, and laboriously wrote out his voluminous notes and manuscripts in so minute a hand that a magnifying glass was sometimes necessary to decipher the script.

Formal résumés of his career are sparse and often inaccurate; he rarely bothered to put himself on record in professional directories. Since he moved from establishment to establishment, state to state, country to country, and job to job, there is little institutional continuity to trace; only his work reflects the inner unity of his life. The host of friends he left behind knew him only in phases and primarily as a teacher, yet his gift for spontaneous intimacy made each feel that he shared with Radin some particular secret.

Radin was a creator and formulator of meaning, a "poet-thinker," a "thinker-artist," a "priest-thinker," and not a man of action, a layman, to use the terms he chose to describe two contrasting, historical, temperamental types—first in *Primitive Man as Philosopher* (1927), then in *Primitive Religion* (1937), and in subsequent works. He could live only in a cerebral "blaze of reality." The degree to which his own commitment to the intellectual life may have led him to see this dichotomy as a universal temperamental division is difficult to determine. Late in life, he adopted the position that these two contrasting temperaments can and usually do complexly coexist in the same person; he seems, therefore, to have invented ideal types, despite his antipathy to such abstract efforts.

In spite of the intensity of his intellectual commitment, Radin was never moved to glorify the position of intellectuals. He regarded them as dependent upon the official academy, which was in turn linked to the ruling establishment, and, along with Franz Boas, he therefore viewed them as bound by convention and, for the most part, incapable of pursuing interests other than their own.

Although he migrated from post to post, Radin was never at a loss for a job. In the words of Julian Steward, "his charm . . . got him about every job in Anthropology in the country. . ." (personal correspondence, 1964). During the early depression

years, when many of his students and colleagues were unemployed, he managed to find support for anthropology from government agencies previously unconcerned with such efforts. For example, his work on Mexican pamphlets in the Sutro Library in San Francisco was supported by the Works Projects Administration. Moreover, he received support over the years from a variety of foundations; he may have offended bureaucrats, but he nevertheless had a talent for attracting patrons.

Radin never conceived of anthropology as merely a specialized discipline; it was for him more a way of life. He was bent on discovering the conditions under which man thrives, and he followed his pursuit where it led him. He objected to professional jargon because it turned the study of man into a mystery. His goal was to attack the "great, recurring, troubling themes in history" (Diamond 1960, p. xviii), to determine basic human nature. "That the cultural pattern hides this knowledge from us forever is a counsel of despair" (1933a, p. 267).

Although Radin felt at home in the cognitive worlds of primitive peoples and admired "the ruthless realism and objectivity" (1953, p. 325) with which they analyzed man, he did not believe in any return to the primitive or entertain any notion of the noble savage. Rather, he believed that the primitive experience is part of our historical consciousness. An understanding of it would pave the way toward an understanding of what is basic to human nature and toward a critical evaluation of civilization.

Radin first studied anthropology in Europe and came to the field of anthropology by way of zoology and history. He became a student primarily of Boas, and it was probably the combined effect of James Harvey Robinson's skeptical humanism and Boas' empiricist insistence on the indivisible potential of primitive and civilized mentalities that originally led Radin to question all notions of primitive inferiority.

In *Primitive Man as Philosopher* (1927) he dismissed theories about the automatism of primitive life as being merely projective of an increasingly routinized modern life. Radin found primitive mentality different in degree but not in kind from civilized mentality. He saw that, despite the variety of cultural forms involved, responses of primitive peoples to the major challenges of life are sophisticated, profound, and—to a civilized person endowed with self-knowledge—understandable ways. Therefore, Radin rejected such theoreticians as the early Lévy-Bruhl and other members of the French sociological school, who stigmatized primi-

tives as being incapable of abstraction, participants in a mystic entity, linguistically inadequate, or lacking individuality. And, although he had some connection with Robert Redfield and the other folk-urban polarity theorists, he was opposed to the view, later adopted by Redfield, that moral consciousness has expanded with civilization. He took great pains to establish moral insights as a primitive characteristic and in general was skeptical of the idea of progress in such areas.

In *The Trickster: A Study in American Indian Mythology* (1955b), for example, he emphasized two themes: first, the ambivalence of the human impulse toward creation and destruction, symbolized in the dual image of the Deity; and second, man's Sisyphean struggle, expressed in trickster mythology, to create, and then constantly to rescue, meaning from a chaos of sensory impressions, biological needs and appetites, enigmas of personal, social, and cosmic origins, and death.

Radin's elaboration of the dual conception of the Deity had descended from Andrew Lang and had been anticipated in the work of Ehrenreich, Boas, Kroeber, and Dixon; indeed both the Americanists and the German historical school had been preoccupied with the issue, but Radin explored the nature of God among primitives more fully than any other American ethnologist. Radin believed that the universal human issues are central to the social and ritual lives of primitive peoples. It was his belief that the relative weight given to certain issues differs for primitive and civilized peoples; this led him to assert that although basic human nature is the same everywhere, it is more visible among primitives.

Primitive life was for Radin closer in its structure and ideology to the roots of comedy and tragedy. His conception that the universal human drama is actually enacted in aboriginal society made his work memorable to nonanthropologists. He attracted such diverse personalities as Mark Van Doren, Lewis Mumford, C. G. Jung, Allen Tate, John Crowe Ransom, John Dewey, and Huntington Cairns.

Radin was constantly aware of the linked exploitative and creative roles of the "priest-thinkers" and thought this link existed to some degree even in primitive societies. He regarded the priest-thinkers as the inventors of religious systems, who thereby reflected their own profound need to create a coherent universe of meaning while catering to the intermittent needs of their laymen followers. He imagined that the priest-thinkers, or shamans, or medicine men, had been the original formulators of the monotheistic synthesis. This synthesis existed



as a social convention of Judaeo-Christian-Islamic civilization, but as a pure faith it is as uncommon among primitives as among ourselves. The drives of the priest-thinkers may have been neurotic and their hunger economic, but for Radin this did not dim the significance of their insights. He was, however, troubled by their exploitative potential.

Radin had a thorough grounding in the history and languages of his own culture. He considered this grounding the basis for his conscious understanding of his own alienation from the prevailing values of his culture, without which "the task of understanding the primitive could not be accomplished."

Radin's general conception of primitive society is the most effective synthesis attempted thus far. He identified three outstanding positive features of aboriginal civilization: respect for the individual, regardless of age or sex; a remarkable degree of social and political integration; and a concept of personal security that transcends all governmental forms and all tribal and group interests and conflicts. His synthesis of primitive economic and social structure, philosophy, religion, and psychology is an implicit, inductive, historical model of primitive society that also serves as a foundation for a historical theory of human nature.

Radin took his first field trip to the Winnebago in 1908 and eventually published monographs and articles on almost every aspect of their lives. In his field work he considered himself a historical reporter and was skeptical of the claims of participant-observers: To avoid the pretentious impressionism common to their approach the observers would have to become members of the tribe, but Radin doubted that any well-qualified ethnologist would be willing to do that. It was not just that he doubted the results of field work done without knowledge of the language but also that he thought field work should be grounded in the lives of specific individuals and not built up from a generalized conception of the individual.

As a linguist Radin was in a class with Sapir and Boas. In addition to the voluminous Winnebago texts, Radin published a series of Wappo (1924), Huave (1929a), and Mixe (1933b) texts, a grammar of Wappo (1929b), notes on Tlapanecan (1933c), and a sketch of Zapotec (1930). He was also concerned with historical linguistics, publishing a classification of Mexican languages (1944) and working for an entire decade on Patwin. His most important linguistic contribution was an early monograph (1919a) on *The Genetic Relationship of the North American Indian Languages*, in which he argued for their essential continental unity. Although Sapir had initially criticized

this idea of unity, he and his student M. Swadesh were later to develop parallel themes on this subject.

R. H. Lowie and Sapir were Radin's closest professional friends, and Radin remained affiliated with the Boas school throughout his life. However, he could not accept their ethnological theories. In his *Method and Theory of Ethnology* (1933a), which, with Lowie's *History of Ethnological Theory* (1937), remains the only systematic work on ethnological theory written by an American, he attacked Boas' quantitative and distributional approach to cultural data. He believed it led to overgeneralized, external, and patchwork histories of traits; in abstractly deduced time perspectives, rather than to specific histories of societies as experienced and created by their members. Radin always felt that there is a contradiction between this aspect of the Americanists' approach and Boas' insistence on prolonged field work and textual analysis within particular societies.

Radin's dispute with Kroeber summarizes his differences with the Americanists. Kroeber's notion of the superorganic and consequent lack of interest in the person in history, his sweeping efforts to classify whole civilizations by configurations of traits and qualities determined by a combination of intuitive and quantitative means, and his insistence on ethnology as a natural science—that is, as having a subject matter composed of discreet, isolable, and objectively determinable elements that can be traced and categorized on their own terms (a view Boas also held)—did violence to Radin's focus on the individual as the locus of culture. Kroeber's view denied Radin's belief that if one probes deeply enough into particular forms, universal meanings will be revealed; Kroeber's view also abused Radin's sense of history as man's agent, as the agency for revealing the nature of man and the necessary conditions for his fulfillment.

Radin, despite the strength of his influence, left no school of students. His final academic affiliation was with Brandeis University, first as professor of anthropology and then as chairman of the department. He had thought that in going to Brandeis he would find a congenial home, but he found the new university no less bureaucratic than the older institutions. He died in 1959 a few days after his heart failed him during a professional lecture in New York City.

STANLEY DIAMOND

[For the historical context of Radin's work, see the biographies of BOAS; KROEBER; LÉVY-BRUHL; REDFIELD; ROBINSON; for discussion of the subsequent development of some of Radin's ideas, see LINGUIS-

TICS, article on HISTORICAL LINGUISTICS; RELIGION; and the biography of SAPIR.]

#### WORKS BY RADIN

- 1906 Zur Netztechnik der südamerikanischen Indianer. *Zeitschrift für Ethnologie* 38:926-938.
- 1913 On Ojibwa Work, 1913. Canada, Geological Survey, Summary Reports [1913]:374 only.
- 1919a *The Genetic Relationship of the North American Indian Languages*. University of California Publications in American Archaeology and Ethnology, Vol. 14, No. 5. Berkeley: Univ. of California Press.
- 1919b The Relationship of Huave and Mixe. *Journal de la Société des Américanistes de Paris* New Series 11: 489-499.
- 1920 *The Sources and Authenticity of the History of the Ancient Mexicans*. University of California Publications in American Archaeology and Ethnology, Vol. 17, No. 1. Berkeley: Univ. of California Press.
- 1924 *Wappo Texts, First Series*. University of California Publications in American Archaeology and Ethnology, Vol. 19, No. 1. Berkeley: Univ. of California Press.
- (1927) 1957 *Primitive Man as Philosopher*. 2d ed., rev. New York: Dover.
- 1929a Huave Texts. *International Journal of American Linguistics* 5:1-56.
- 1929b *A Grammar of the Wappo Language*. University of California Publications in American Archaeology and Ethnology, Vol. 27. Berkeley: Univ. of California Press.
- 1930 A Preliminary Sketch of the Zapotec Language. *Language* 6:64-85.
- 1931 Concept of Right and Wrong. Pages 818-827 in Victor F. Calverton (editor), *The Making of Man: An Outline of Anthropology*. New York: Modern Library.
- 1933a *The Method and Theory of Ethnology: An Essay in Criticism*. New York: McGraw-Hill.
- 1933b Mixe Texts. *Journal de la Société des Américanistes de Paris* New Series 25:41-64.
- 1933c Notes on the Tlappanecan Language of Guerrero. *International Journal of American Linguistics* 8:45-72.
- 1934 *The Racial Myth*. New York: McGraw-Hill.
- (1937) 1957 *Primitive Religion: Its Nature and Origin*. New York: Dover.
- 1944 The Classification of the Languages of Mexico. *Tlalocan* 1:259-265.
- 1945 *The Road of Life and Death: A Ritual Drama of the American Indians*. Bollingen Series, No. 5. New York: Pantheon.
- 1953 *The World of Primitive Man*. New York: Schuman. → A paperback edition was published in 1960 by Grove.
- 1955a The Literature of Primitive Peoples. *Diogenes* 12: 1-28.
- (1955b) 1956 *The Trickster: A Study in American Indian Mythology*. New York: Philosophical Library.

#### WORKS TRANSLATED OR EDITED BY RADIN

- ADLER, ALFRED (1909-1920) 1964 *The Practice and Theory of Individual Psychology*. Translated by Paul Radin. Rev. ed. New York: Harcourt. → First published as *Praxis und Theorie der Individual-Psychologie*. Contains 28 papers originally published in medical journals between 1909 and 1920.
- African Folktales and Sculpture*. (1952) 1965 2d ed. Folk tales selected, edited and introduced by Paul Radin. London: Secker & Warburg.

[BLOWSNAKE, SAM] (1920) 1963 *The Autobiography of a Winnebago Indian*. Edited and translated by Paul Radin. New York: Dover. → The writer is referred to throughout the notes as S. B. The autobiography proper closes with Part I. Part II embodies the system of instruction used among the Winnebago and forms a unit in itself. Part I was also published as *Crashing Thunder: The Auto-biography of an American Indian*. CALIFORNIA STATE LIBRARY, SACRAMENTO, SUTRO BRANCH, SAN FRANCISCO 1939-1940 *Catalogue of Mexican Pamphlets in the Suto Collection: (1623-1888)*. Edited by Paul Radin. 2 vols. San Francisco: Mimeographed.

#### SUPPLEMENTARY BIBLIOGRAPHY

- BOAS, FRANZ (1887-1936) 1955 *Race, Language and Culture*. New York: Macmillan.
- DEWEY, JOHN (1927) 1957 Foreword. In Paul Radin, *Primitive Man as Philosopher*. 2d ed., rev. New York: Dover.
- DIAMOND, STANLEY (editor) 1960 *Culture in History: Essays in Honor of Paul Radin*. New York: Columbia Univ. Press. → A bibliography of Radin's writings appears on pages 1001-1010.
- DUBOIS, CORA 1960 Paul Radin: An Appreciation. Pages ix-xvi in Stanley Diamond (editor), *Culture in History: Essays in Honor of Paul Radin*. New York: Columbia Univ. Press.
- EHRENREICH, PAUL 1906 Götter und Heilbringer: Eine ethnologische Kritik. *Zeitschrift für Ethnologie* 38: 536-610.
- HYMES, DELL H. 1965 Some North Pacific Coast Poems: A Problem in Anthropological Philology. *American Anthropologist* New Series 67:316-341.
- LOWIE, ROBERT H. 1937 *The History of Ethnological Theory*. New York: Farrar & Rinehart.
- SCHMIDT, WILHELM 1919-1920 Die kulturhistorische Methode und die nordamerikanische Ethnologie. *Anthropos* 14/15:546-563.
- SCHMIDT, WILHELM 1921-1922 Die Abwendung vom Evolutionismus und die Hinwendung zum Historizismus in der Amerikanistik. *Anthropos* 16/17:487-519.
- SCHMIDT, WILHELM 1926-1955 *Der Ursprung der Gottesidee: Eine historisch-kritische und positive Studie*. 12 vols. Münster (Germany): Aschendorff.
- VAN DOREN, MARK 1945 Foreword. In Paul Radin, *The Road of Life and Death: A Ritual Drama of the American Indians*. New York: Pantheon.

#### RADIO

See COMMUNICATION, MASS.

#### RADLOV, VASILII, and BARTOLD, VASILII

Vasilii Radlov, also known as Wilhelm Radloff (1837-1918), was a Russian Turkologist, museum director, explorer, and teacher. His scholarly contributions covered the ethnography, history, folklore, text redactions, archeology, and material culture of the Turkic peoples of central Asia and southern Siberia, and the linguistics, including the comparative lexicography, of all Turkic peoples. In addition, he was a professor, an academician, and a government adviser on Turkic matters.



Vasilii V. Bartol'd, also known as Wilhelm Barthold (1869–1930), academician and professor at St. Petersburg, specialized in the history of Islam and of the Turkic peoples and Iranians (Tadjiks) of central Asia. He combined a life of scholarship and teaching with long and frequent field trips. Bartol'd's work is being considered here together with Radlov's, since it contained important modifications of Radlov's theories.

Of German origin, Radlov was introduced to Oriental studies at the University of Berlin in the 1850s by two of Wilhelm von Humboldt's disciples, A. F. Pott and H. Steinthal. These men were engaged in a struggle against Gobineau's doctrine of the inequality of races, and Radlov also eventually devoted much of his energy to affirming the like potentialities of all human groups. He discounted the influence of biological race on human development and, on the contrary, echoed Humboldt's older ideas: while intelligent occupation does develop the critical faculties, nevertheless sound judgment and sharpness of wit may be found among peoples without letters or education as, for example, among the peoples of central Asia. These ideas of Humboldt's constitute a sort of "noble barbarian" theory.

On completing his studies, Radlov taught German and Latin for 12 years at a Gymnasium in Barnaul, Siberia. This period was essential to his future career, for it gave him close and uninterrupted access to the Turkic peoples of the Altai and Sayan mountains, of southern Siberia, and of Kazakhstan and Kirgizia. His ethnographic descriptions, text collections, and linguistic analyses were begun at this time.

Following his Siberian period, he went to St. Petersburg. Here he published a series of volumes, *Die alttürkischen Inschriften der Mongolei* (1894–1897), in which Turkic script carved on stelae was deciphered, transliterated, and translated. N. Yadrintsev, Radlov himself, and others had discovered these inscriptions in the region of the upper Yenisei and Orkhon rivers in western Mongolia; Radlov and a Danish contemporary, V. Thomsen, were able to master the script and language. Bartol'd, then a junior colleague of Radlov at the University of St. Petersburg, wrote a chapter in the series on the historical meaning of the inscriptions. Radlov's contributions lie at the foundation of all subsequent work on this civilization; his interpretations, modified and extended by later work, stand to this day. He identified the documents as those of a pastoral Turkic people who built an indigenous state of great extent but simple form in the sixth to eighth centuries A.D. The script

was their own; it had a superficial resemblance to Scandinavian runes, and both were connected ultimately, through diffusion, with alphabetic systems of the ancient Near East. Many of the inscriptions were multilingual, and the inclusion of Chinese translations of the Turkic shows that this civilization had connections with east Asia as well as with the West.

At this time Radlov also translated (1891–1910) the *Kudatku* (better, *Kutadgu*) *Bilik*, a lengthy moral-didactic poem of the medieval Uighurs, Turkic-speaking peoples of central Asia, a poem composed by the royal adviser, Yūsuf Khāss-hājib of Balasagun. In a preface to the poem, Radlov provided both an account of Uighur society and a general theory of state formation by usurpation, with particular reference to the Uighurs. This was a theory to which he repeatedly adverted and which was further developed by Bartol'd. According to Radlov, pastoral nomadic tribes developed a permanent political state of their own only when a powerful or wealthy person usurped the popular authority. The usurper's group was then eponymously named, whereby the leader sought to stabilize his power within his descent line. Radlov based this theory of the origin of the state upon his ethnographic observations of just such a process among the Kazakhs; subsequently he extrapolated this process to ancient times. The method of applying current ethnographic observation to reconstructions of bygone eras and peoples was used by L. H. Morgan and other theorists of evolutionary stages, especially in the nineteenth century.

Radlov worked out a three-stage theory of evolution in his book *Aus Sibirien* (1884), a general ethnography of central and northern Asia. He believed that human life in this area developed and made the transition from a hunting to a pastoral nomadic stage and then to agriculture. Each stage was characterized not only by the economy associated with its name but also by a particular social and political organization and a predominant religion. For the pastoral Turkic peoples of central Asia the predominant religion was shamanism. At the same time Radlov noted that the diffusion of Christianity through contact with Russians, the diffusion of Islam through contact with Iranian and Semitic peoples, and the diffusion of Buddhism from India, Tibet, and the Mongols all changed the indigenous religion. The Russian Orthodox church regarded shamanism as evidence of unclean powers at work among the peoples of Asiatic Russia. Radlov controverted this view, asserting that shamanism was a religion to be studied in its own

right. He succeeded to a degree in modifying the official Russian view of shamanism, but his educational ventures among the Turkic peoples were opposed by those who supported education only when it was connected with the propagation of Christianity.

His evolutionary views shaped the policies he advocated for the progress of the Turkic peoples. He saw contemporary nomadism as a retarding force. The pastoral empires of Asian history were brilliant developments of the past but could not meet the needs of the peoples in modern times. Radlov's analyses of the ancient and medieval Turkic khanates convinced him that they were essentially autocratic institutions which prevented the Turkic pastoralists from sharing in the progressive changes occurring in European life. This view is the dominant one in Soviet policy regarding the area today.

Radlov was active also in collecting the folklore of the Turkic peoples of southern Siberia and central Asia (1866-1907). The material he collected was published (in part) in ten volumes, the first systematic and precise coverage of this ethnographic field. Moreover, he brought out a monumental comparative lexicon of the Turkic dialects and languages. The study of the Turkic peoples was transformed into a scientific undertaking by virtue of these efforts. Radlov's studies were related to the broad ethnographic, historical, lexicographic, and evolutionary problems of his time.

His activities in public affairs were as impressive as his scholarly contributions. He advanced an enlightened policy toward the peoples of Asia; he gave material and spiritual support to exiled revolutionary scholars; and he was a vital force in transforming the St. Petersburg Kunstkammer of Peter the Great into the modern Museum of Anthropology and Ethnography.

Bartol'd's work was narrower in focus and less theoretical in tone than that of Radlov. He was concerned with the interaction between an individual and society, more particularly with refining Radlov's theory of state formation by usurpation; he examined the role of Chinghiz Khan and of the founders of other pastoral empires of Asia from this point of view. Bartol'd eschewed inferential reconstruction of the past and committed himself straightforwardly to cultural and social history; thus, his interpretations of the personality of Chinghiz Khan are closely related to the textual record.

In line with this historical bent, Bartol'd studied the literate peoples of central Asia and their written record, he treated the nonliterate cultures mar-

ginally, and references to them appear only occasionally in his writings. Nevertheless, he did criticize Radlov's designation of shamanism as the religion of the Turkic pastoral nomads. Bartol'd associated shamanism with the hunting cultures of the polar regions, although he agreed that it is found among those pastoralists living in remote areas such as the Altai Mountains. Bartol'd also criticized Radlov for failing to treat the function of shamanist religion in society; such an approach would have revealed that the role played by shamanism in hunting cultures was even greater than it was among pastoralists.

Bartol'd combined in his studies the results of both library and field work and alternated between specialized histories, such as those of cotton-growing in central Asia or of Christianity in Asia, and broader syncretistic topics, such as the cultural history of Turkestan, of Iran, and of the Turkic peoples of central Asia. He also compiled a bibliographic survey *La découverte de l'Asie* (1911). His brief and brilliant conspectus of Turko-Mongol history, published in Tashkent in 1929, reduced to a short essay the political dynamics common to all central and inner Asia ([1893-1929] 1956-1962, vol. 3, pp. 73-170). Of particular importance to Western scholarship are his short characterizations of Turkic peoples—Tadjiks, Kazakhs, Kirgiz, Turkmens, Kalmuks, and other peoples of Asia and the Caucasus—in the *Encyclopaedia of Islam* (1913-1936).

LAWRENCE KRADER

[For the polemical context of Radlov and Bartol'd's work, see the biography of GOBINEAU. The biography of BOAS and the article on BOGORAZ, STERNBERG, and JOCHELSON contain relevant background material. For discussion of the subsequent development of their ideas, see PASTORALISM.]

#### WORKS BY RADLOV

- 1866-1907 RADLOV, VASILII V. (compiler and translator), *Proben der Volkslitteratur der türkischen Stämme*. 10 vols. in 18. St. Petersburg: Kaiserliche Akademie der Wissenschaften.
- (1884) 1893 *Aus Sibirien: Lose Blätter aus meinem Tagebuche*. 2 vols. in 1. Leipzig: Weigel.
- 1891-1910 YÜSUF, KHÄSS-HÄJIB *Das Kudatku Bilik des Jusuf Chass-Hadschib aus Bälasagun*. Edited by Wilhelm Radloff. 2 vols. St. Petersburg: Kaiserliche Akademie der Wissenschaften.
- (1893-1911) 1963 *Opyt slovaria türkskikh narechii: Versuch eines Wörterbuches der Türk-dialekte*. 4 vols. Moscow: Izdatel'stvo Vostochnoi Literatury.
- 1894-1897 *Die alttürkischen Inschriften der Mongolei*. St. Petersburg: Kaiserliche Akademie der Wissenschaften. → Issued in four parts. Part 4 contains a chapter by Bartol'd, "Die historische Bedeutung der alttürkischen Inschriften."



## WORKS BY BARTOL'D

- (1893–1929) 1956–1962 *Four Studies on the History of Central Asia*. 3 vols. Leiden (Netherlands): Brill. → First published in Russian. Volume 1: *A Short History of Turkestan* (1922) 1956; *History of the Semirechye* (1893) 1956. Volume 2: *Ulugh-beg* (1918) 1958. Volume 3: *Mir 'Al-shir* (1928) 1962; *A History of the Turkman People* (1929) 1962.
- (1900) 1958 *Turkestan Down to the Mongol Invasion*. 2d ed., enl. & rev. London: Luzac. → First published in Russian.
- (1911) 1947 *La découverte de l'Asie: Histoire de l'orientalisme en Europe et en Russie*. Paris: Payot. → First published as *Istoriia izucheniia vostoka v Evrope i Rossii*.
- 1935 12 *Vorlesungen über die Geschichte der Türken Mittelasiens*. Berlin: Deutsche Gesellschaft für Islamkunde.
- Sochineniia*. Vols. 1—. Moscow: Izdatel'stvo Vostochnoi Literatury, 1963—. → The first three volumes of this projected ten-volume edition of Bartol'd's collected works have appeared.

## SUPPLEMENTARY BIBLIOGRAPHY

- Encyclopaedia of Islam*. 4 vols. 1913–1936 Leiden (Netherlands): Brill. → Bartol'd contributed over 200 articles.
- KRADER, LAWRENCE 1958 *Feudalism and the Tartar Polity of the Middle Ages. Comparative Studies in Society and History* 1:76–99.
- SAVICKIJ, P. 1931 V. V. Bartol'd: Une page de l'histoire de la conscience nationale russe. *Monde slave* [1931]: no. 8:273–288.

## RAE, JOHN

John Rae (1796–1872) was born in Scotland, but it was while he was living in Canada that he wrote his most important book, a work on the theory of capital and of economic development. He migrated to Canada in 1822, having acquired at the universities of Edinburgh and Aberdeen a knowledge not only of the classics but also of mathematics and, to some degree, of medicine. In Canada he held several schoolmasterships, from the last of which he was discharged in 1848, in part because he had become involved in religious and educational controversy. In financial straits and with poor professional prospects, he traveled across Panama to California, where he remained until 1851, and then went to the Hawaiian Islands, where he hoped to regain his health and resume a professional career. He supported himself at first by teaching, later mainly through medical and judicial posts. In 1871, ill and without funds, he returned to New York to live with a wealthy friend. He died within a year.

The treatise on which Rae's fame mainly rests, *Statement of Some New Principles on the Subject of Political Economy Exposing the Fallacies of the*

*System of Free Trade and of Some Other Doctrines Maintained in the Wealth of Nations*, was published in Boston in 1834. The full title is revealing; the book was a by-product of a never completed study of the Canadian economy and was essentially an argument with Adam Smith.

Rae's *New Principles* has been called by Schumpeter "a theory of capital, conceived in unprecedented depth and breadth," containing, as it does, two essential components of the theory that Böhm-Bawerk was to erect: "the proposition that 'lengthening' the process of production (postponement) will usually increase the physical amount of final product . . . and the proposition that 'the actual presence of the immediate object of desire' will give to it, in our valuation, a decisive advantage over an exactly similar object that is expected to become available at some future date, even if this expectation be perfectly certain" (Schumpeter 1954, p. 469).

Apart from his important contribution to the theory of capital, what distinguishes Rae's political economy most strikingly from that of his orthodox contemporaries is his treatment of the neglected subject of invention (with which he was familiar at first hand) and his discussion of the capacity of legislators to foster invention and capital formation by preventing dissipative luxury, by promoting education, morality, and science, by maintaining order, and by affording protection to infant industries with real potential. His emphasis thus reflected his experience in a frontier society, as did his recognition of the role of banks in promoting growth and of the significance of durability and yield for capital allocation; his frontier experience may also have prompted his Veblen-like assessment of luxury.

Rae's boyhood ambition, according to his later reminiscences, had been a large-scale study of natural history, especially the natural history of man, of the social attitudes that condition man's behavior, and of changes in these attitudes. He continued all his life to do "research" on this subject, and although he never produced a major work, he did write a number of papers in the general area that we would now call social science. While in Hawaii he wrote an article on the Polynesian language, anticipated Wundt's ideas on the gesture theory of language, and transcribed a Hawaiian legend. In his *New Principles* he had developed the notion of the "effective desire of accumulation"—a phrase that was adopted by John Stuart Mill, who quoted much from Rae (see Schumpeter 1954, p. 469)—and in correspondence with Mill, Rae developed this idea further,

emphasizing the role of social "sentiments" in the formation of capital. In his letters to Mill he attributed the decline of Hawaii's population to a weakening of the "effective desire for offspring," which in turn weakened the "effective desire of accumulation."

The *New Principles*, as R. Warren James (1951) has noted, had rather limited influence in the nineteenth century, despite its support from Nassau Senior and, above all, from Mill; its translation into Italian; and its generally favorable reception in the United States and Canada. Reasons for this limited influence may be the small number of copies printed, Rae's residence in a then poor and undeveloped country, and the unpopularity of capital theory in England. The articles on Rae by C. W. Mixter (1897; 1902) and Mixter's new edition of the *New Principles*, somewhat rearranged, annotated, and titled *The Sociological Theory of Capital*, made Rae's work known to a distinguished circle of scholars, among them Böhm-Bawerk, Irving Fisher, and Knut Wicksell. Recently there has appeared *John Rae, Political Economist: An Account of His Life and a Compilation of His Main Writings*; Volume 1 contains a splendid biography by the editor, R. Warren James, which has finally made Rae's position in the history of economic analysis secure.

JOSEPH J. SPENGLER

[For the historical context of Rae's work, see the biography of SMITH, ADAM; for discussion of the subsequent development of Rae's ideas, see CAPITAL and the biographies of BÖHM-BAWERK; FISHER, IRVING; MILL, article on ECONOMIC CONTRIBUTIONS; WICKSELL; WUNDT.]

#### WORKS BY RAE

(1834) 1905 *The Sociological Theory of Capital*. New ed., annotated by C. W. Mixter. New York: Macmillan. → First published as *Statement of Some New Principles on the Subject of Political Economy Exposing the Fallacies of the System of Free Trade and of Some Other Doctrines Maintained in the Wealth of Nations*.

*John Rae, Political Economist: An Account of His Life and a Compilation of His Main Writings*. Compiled by R. Warren James. 2 vols. Univ. of Toronto Press, 1965. → Volume 1: *Life and Miscellaneous Writings*. Volume 2: *Statement of Some New Principles on the Subject of Political Economy*. Volume 1 contains an excellent biography by the editor, R. Warren James.

#### SUPPLEMENTARY BIBLIOGRAPHY

GOODWIN, CRAUFURD D. W. 1961 *Canadian Economic Thought*. Durham, N.C.: Duke Univ. Press.  
JAMES, R. WARREN 1951 *The Life and Work of John Rae*. *Canadian Journal of Economics and Political Science* 17:141-163.

LEHMANN, HELMUT 1937 *John Raes Werk: Seine philosophischen und methodologischen Grundlagen*. Dresden (Germany): Dittert.  
MIXTER, CHARLES W. 1897 A Forerunner of Böhm-Bawerk. *Quarterly Journal of Economics* 11:161-190.  
MIXTER, CHARLES W. 1902 Böhm-Bawerk on Rae. *Quarterly Journal of Economics* 16:385-412.  
SCHUMPETER, JOSEPH A. (1954) 1960 *History of Economic Analysis*. Edited by E. B. Schumpeter. New York: Oxford Univ. Press.  
SPENGLER, JOSEPH J. 1959 John Rae on Economic Development: A Note. *Quarterly Journal of Economics* 73:393-406.

## RANDOM NUMBERS

Random numbers are numbers generated by a mechanism that produces irregularity, in a sense that will be made precise. There are four major uses of random numbers: (1) To protect against selective bias in the acquisition of information from sample surveys and experiments. In these same contexts, random numbers provide a known probability structure for statistical calculations. (2) To gain insight, by simulation, into the behavior of complex mechanisms or models. (3) To study theoretical properties of statistical procedures, such as efficiency of estimation and power of statistical tests. (4) To obtain approximate solutions to other mathematical problems. In each of these applications, random numbers are used to simulate observations from an arbitrary probability distribution. It is often useful to think of random numbers as an idealization of the successive outcomes of a carefully made gambling device, say a roulette wheel.

**Definitions.** A definition of random numbers is a specification of the probabilistic properties of the process that generates them, rather than a specification of the properties of the numbers themselves. One can only partially test and verify the numbers themselves; they are only specific outcomes of the process. To clarify the distinction between a process and an outcome of the process, consider the example of coin tossing. In stating that the probability of getting a head in flipping an honest coin is  $\frac{1}{2}$ , one implies that before the coin is flipped the process of flipping gives probability  $\frac{1}{2}$  of observing a head. Once the coin is flipped, however, either a head or a tail shows; the situation is determined and not probabilistic.

Two concepts are involved in specifying the process generating random numbers: independence and a specific type of nondeterminism. In terms of the coin-flipping process, independence implies that the process that flips the coin is not



influenced by prior flips and nondeterminism implies that one cannot state with certainty whether a head or a tail will occur on a flip. In the case of an honest coin one can only say that the probability is  $\frac{1}{2}$  that a head will occur.

A process is said to generate random numbers in a strict, or theoretical, sense when the process provides random variables that are (1) independent in the sense of probability and (2) governed by the same continuous uniform distribution. Independence implies that knowledge of some numbers generated by the process cannot help in predicting other numbers. The uniform distribution requirement ensures that the numbers generated lie within an interval; the probability that a random number lies in a subinterval is given by the length of the subinterval divided by the length of the entire interval [see DISTRIBUTIONS, STATISTICAL; PROBABILITY].

A process generates random (decimal) digits when it provides random variables that are independent and take the values 0, 1, ..., 9 with equal probability  $\frac{1}{10}$ , that is, the random variables follow a discrete uniform distribution. Similar definitions apply to random digits for bases other than 10, for any set of contiguous integers, or, in general, for any set following a discrete uniform distribution.

No real process can truly generate random numbers in the strict sense, since real processes must always generate rational numbers. The term "random numbers" is, however, often applied to processes that generate independent, equispaced, equally probable rational numbers when those numbers are so numerous that the discrete process may be considered an approximation to a continuous one. It is necessary, nonetheless, to keep this distinction between theory and practice in mind. Similarly the terms "random digits" and "random numbers" are in practice applied to real processes that presumably can never provide values that are *exactly* independent or uniform. Further, the term "random numbers" is often loosely used to apply to any of the above possibilities. Finally, these terms are even sometimes used to mean any number-producing process—probabilistic or not—where the produced values appear approximately equally often and where there is no evident pattern in the series of outcomes. See Hammersley and Handscomb (1964), Shreider (1962), or Young (1962) for additional discussion. This terminological thicket is not so tangled as might appear at first, since the meaning intended is usually clear from context.

**Uses of random numbers.** Random numbers, digits, etc. permit the construction of probability samples, which in turn permit the use of present-day statistical inference. In sample surveys, units must be selected from the sampled population by probability sampling if statistical analysis, in the ordinary sense, is properly to be done. The simplest case is that in which each unit of the population has an identification number (from 1 to  $N$ , say). Random integers (in the interval 1 to  $N$ ) are chosen and the corresponding units form the sample. In practice, sampling without replacement is usually used, and here random permutations are needed; these are discussed below. Also in practice, sampling designs of various degrees of complexity are common, using devices of stratification, clustering, etc. [see SAMPLE SURVEYS].

In a similar way, random integers or permutations are used in experimental design for deciding which experimental treatments are to be applied to the various physical units: mice, men, fields, etc. In both these cases, the use of a random device serves to protect against selective bias, for example, assigning the healthiest subjects to one experimental therapy. Of equal importance, however, is the fact that the use of a random device provides a firm foundation for statistical inference [see EXPERIMENTAL DESIGN].

Random numbers can also be very useful in studying models of complex physical mechanisms or abstract theories, whenever a model involves variables with probabilistic (stochastic) elements from arbitrary probability distributions. For example, in models of fatigue or of learning one can simulate the performance of many individuals by generating observations from a hypothetical probability distribution depicting performance and then using the generated results to estimate actual performance. In essence, one is sampling from the hypothetical population to estimate how the "real" population can be expected to perform. Random numbers are used to obtain observations from an arbitrary probability distribution function. If observations  $\{x\}$  are desired from a probability distribution  $F(x)$ , and if  $\{u\}$  denotes values of random numbers in the unit interval  $(0,1)$ , that is, observations from the unit uniform distribution, then an observation value  $x$  from  $F(x)$  is obtained by generating a value for  $u$  and solving for  $x$  in the relationship  $x = F^{-1}(u)$ . For the derivation and conditions necessary for this result, see, for example, Mood (1950, p. 107). More economical ways of obtaining a conversion of random numbers to observational values of  $F(x)$  can often be achieved

For example, several methods exist for obtaining observations from the normal distribution function (see Box & Muller 1958; Muller 1958; Teichroew 1965; Hull & Dobell 1962).

The use of random numbers has also been considered for solving physical or mathematical problems that appear to be completely nonprobabilistic. For example, assume that a mathematical function is given and it is desired to find (by integration) the area under its curve between specified limits of the independent variable. The solution to this problem can be approximated by use of random numbers. Imagine, for convenience, that the area to be estimated lies within the unit square; then by generating pairs of random numbers that lie in the unit square and comparing the number that lie under the curve to the total number of pairs generated, one obtains an estimate of the area under the curve. The general basis for this approach is considered, for example, by George W. Brown (1956, p. 283).

With the advent of modern digital computers this approach to mathematical problems has been given considerable attention and is referred to as the Monte Carlo (sampling) approach (see Florida, Univ. of 1956; Hammersley & Handscomb 1964; Shreider 1962). Indeed, the Monte Carlo approach can be used to explore theoretical questions of statistics that are not readily solved by direct analysis (see Teichroew 1965). Without going into detail, it can be stated that random numbers can also be used to mix strategies in complex decision processes, an application of growing usefulness [see DECISION THEORY].

**Sources of random numbers.** Physical and mathematical processes have both been employed to generate random numbers. It is very difficult to determine whether these sources provide satisfactory numbers; the closeness of agreement between the real-world process and the hypothetical model has been measured by rather general purpose statistical tests, which at best can give only partial answers.

Until recently, the major sources of random numbers have been published tables derived primarily by using physical devices, some of which have been improved with the aid of mathematical processing. A list of prepared tables of random numbers is presented below.

The physical processes used include (1) drawing capsules from a bowl, with replacement and remixing after each drawing; (2) flashing a beam of light at irregular time intervals on a sectioned rotating disk, so as to obtain section numbers;

(3) recording digitized electronic pulses whose time distribution of occurrences is supposedly random; and (4) counting the number of output pulses generated by the radioactive decay of cobalt 60 during constant time intervals.

Physical processes have several disadvantages as sources of random numbers. Unless one stores the numbers generated by a physical process one cannot have them available and cannot regenerate the specific sequence used if it becomes necessary to check or repeat a computation. Furthermore, the early types of physical processes were difficult to maintain and use. Interest has shifted to the use of mathematical processes that can be programmed for digital computers, since large tables cannot as yet be economically entered and stored in the high-speed memories of computers. Most such mathematical processes center about a recurrence relation, where the next number of a sequence is derived from one or more prior numbers of the sequence.

These mathematical processes are deterministic, that is, any number of a sequence is a specified function of prior numbers of the sequence. For this reason, one could easily raise objections to the use of these processes and prefer the use of physical processes. However, as will be noted below, there is a strong pragmatic justification for these mathematical procedures.

Several methods use the output of a physical process as initial "random" numbers, which are improved by use of mathematical relationships such as those provided by Horton and Smith (1949). These techniques have been called compound randomization techniques, and they appear to hold much promise.

**History.** Tables of random numbers and random digits are a relatively recent development, having been available only since 1927. This is surprising, considering that, almost a century earlier, mechanical methods and shuffling devices for random numbers were already employed to examine probability problems. Samples for studying theoretical properties of various statistics had been reported in 1908 by "Student" (W. S. Gosset), who investigated the distribution of the standard deviation and the ratio of the mean to the standard deviation of samples by shuffling and drawing at "random" from a population of 3,000 cards (see Gosset 1907-1937). This study and earlier sampling studies using dice to estimate correlations among events—for example, Darbishire in 1907 and Weldon in 1906—are reviewed in Teichroew (1965).



Even after it was recognized that random numbers would provide a reliable and economical tool in selecting samples, a satisfactory method to generate them still remained to be found. Karl Pearson was instrumental in encouraging L. H. C. Tippett to develop a table of random digits. In the introduction to that table, Pearson mentioned that Tippett found that drawing numbered cards from a bag was unsatisfactory and that better results were obtained by using digits selected from a 1925 census report. With this approach, Tippett (1927) provided a table of 41,600 digits. Subsequently several analyses of this table were made and larger tables using other methods have since appeared. With the introduction of punched-card equipment, calculators, and digital computers, other methods of creating random numbers have been considered. Interest has focused on additional ways to test for the adequacy of numbers produced by various methods. Recognizing that in long sequences of random numbers, one might observe patterns of numbers that might be objectionable for a specific use, Kendall and Smith (1938, p. 153) provided criteria of local randomness by which to judge whether a set of numbers is acceptable for a specific application. One could raise objections to the use of the concept of local randomness to include or exclude specific sequences of numbers because each has one or more sets of subsequences that do or do not occur in some desired order. The practical and philosophical questions here are not solved and they raise doubt in some minds about the very foundations of probability theory. For example, G. Spencer Brown (1957) has given considerable discussion to the apparent paradox resulting from the exclusion of specific patterns of numbers.

The first of the mathematical generation processes, which is discussed below, appears to have been suggested by John von Neumann, who proposed the middle square method; this was followed by a multiplicative congruence method by Lehmer, which appeared in 1951.

In passing it should be noted that although Fisher did not directly contribute to the development of techniques for random numbers until 1938, when he and Yates provided their tables, his influence in suggesting the use of randomization in experimental design undoubtedly stimulated much interest in the subject.

**Tables of random digits.** Since the publication of the table of 41,600 digits by Tippett in 1927, several major tables have appeared. In 1938-1939 two were published: the first, by Fisher and Yates (1938), contains 15,000 digits based on entries in a table of logarithms; the second, by Kendall and

Smith (1938; 1939), contains 100,000 digits based on readings from a rotating disk. In conjunction with the publication of their table, Kendall and Smith introduced their four tests of randomness and the concept of local randomness. Several workers have examined these three tables. In 1953, Good raised the question of the need for a modification of one of the tests by Kendall and Smith. From 1939 through the middle of the 1950s several other tables were developed. The most ambitious and largest was the one developed by workers at the RAND Corporation, using the output of an electronic pulse generator and a compound randomization technique to "improve the numbers" (see Brown 1951). These tables originally had limited distribution and were also available on punched cards. In 1955, a table appeared as a book under the title *A Million Random Digits With 100,000 Normal Deviates* (RAND Corp. 1955). This book is now more or less the standard source. The table of normal deviates was derived by converting random numbers to observations from the unit normal distribution by the method of inversion mentioned earlier. Several earlier workers had also published tables of normal deviates based upon, for example, Tippett's table. See Section 13 of Greenwood and Hartley (1962) for an extensive list of tables on random digits and random observations from other distributions.

Each publication of random digits contains, in addition, instructions for the use of the tables. There are unresolved questions about the consequences of the repeated use of the same table.

As previously mentioned, mathematical methods for producing approximately random numbers were introduced for reasons of convenience and efficiency. When a sequence of random numbers is used in a computation, it is often desirable to be able to reproduce the sequence exactly so that comparisons can be made among related computations that use the same sequence. The mathematical generation processes that have been provided to date all satisfy the requirement of reproducibility; that is, given a definite starting point in the process, the same sequence of subsequent numbers can be obtained in repeated applications of the process. The mathematical processes that have been introduced reflect the fact that the electromechanical calculators and early digital computers had very limited memories and generally could perform only one arithmetical or logical operation at a time. It was, therefore, not economical to store large volumes of random numbers, and the generation processes, to conserve memory and computer time, usually used at each step only

one or two prior numbers of the process, which were combined by simple computer operations. It remains to be seen whether more involved generation processes will be introduced to take fuller advantage of advances in the speed and capacity of computers.

**Mathematical generation processes.** A mathematical process for obtaining sequences of numbers in a deterministic manner cannot generate sequences of truly random numbers. Despite this limitation, however, deterministic processes have a practical advantage over random processes, in that a random process can generate very undesirable, although unlikely, sequences of numbers, which would provide questionable samples or experimental designs. One can, on the other hand, sometimes design a deterministic process that generates numbers in a sufficiently haphazard arrangement to satisfy criteria of approximate randomness essential to an application. Processes can be designed so as to give special importance or weight to specific sequences of numbers. Such weighting techniques are similar to stratification techniques used in sampling applications; these ideas are considered, for example, by George W. Brown (1956), Daniel Teichroew (1965), and H. A. Meyer (Florida, Univ. of 1956). Hammersley and Handscomb (1964, p. 27) use the term "quasi random numbers" when it is known that a sequence of numbers is nonrandom but possesses particular desirable statistical properties.

To prevent misunderstanding about the interpretation of numbers generated by a deterministic mathematical process, the process will be referred to as one that generates pseudo random numbers when the numbers are produced with the intent to use them as if they were random numbers. Use of the term "pseudo random numbers" relates to the intent rather than the performance of the process, and the suitability of a process that generates pseudo random numbers must be judged relative to each specific application. For example, assume that in a specific application it is necessary to be able to generate even and odd numbers with the same probability. Further, assume there exists some method for generating statistically independent numbers, with  $r$  as the probability that a number generated will be odd and  $q = 1 - r$  as the probability that it will be even. It is desired that  $r$  equal  $q$ . The following device can be employed to satisfy the requirement: Generate and use the numbers in pairs. If the pair is odd-odd (with probability  $r^2$ ) or even-even (with probability  $q^2$ ), ignore the pair. If the pair is odd-even, call it an odd occurrence; if the pair is even-odd,

call it an even occurrence. Note that the probabilities of an odd occurrence or an even occurrence are now equal; namely,  $r(1 - r) = (1 - r)r$  so that the conditional probabilities are both  $\frac{1}{2}$ .

If the generation process is being performed to obtain  $p$ -position numbers in a given number system (the number 725 is a three-position decimal number), then the largest number of distinct numbers possible is the base of the number system raised to the  $p$ th power. For example, in the decimal base the limit is  $10^p$ ; thus the number of distinct numbers that can be specified using three positions is limited to 1,000. Although it follows that the number of possible distinct pseudo random numbers is limited by the number system and the number of positions used, this limitation is not necessarily serious. Some of the numbers will be repeated whenever a sufficiently long sequence of  $p$ -position numbers is generated. If the sequence being generated begins to consist of numbers in a cycle pattern that will be repeated without interruption, then the length of this cycle is called the period of the process. The early generation methods were designed to provide sequences with a period as large as possible for a given  $p$ . This was partially motivated by the belief that it might help ensure adequate haphazardness of the numbers. The period of a specific process can be either smaller or larger than the number of distinct  $p$ -position numbers that can be generated. The determination of the period, if any, of a process can be very difficult and can depend on how many prior numbers are used in determining the next number to be generated. It is often felt that it is important for the period to be longer than the number of pseudo random numbers needed for a specific application or experiment.

In addition to seeking a long period, it is customary to be sure that a process does not contain the same number too often. Some processes also need to be checked to ensure that they do not degenerate, that is, begin to repeat a single number pattern with a short period. In spite of the deterministic nature of a process, it may be very difficult to analyze its behavior except by subjecting actual outcomes to statistical tests. For example, the behavior of the middle square method, which was proposed by von Neumann, has been difficult to analyze. The method proposed by Lehmer, however, is much more tractable to analysis by number theory, and it has features in common with many other methods proposed by subsequent workers, such as the multiplicative congruence method described below. Hull and Dobell (1962) have compiled a very comprehensive survey of all



mathematical methods and studies made through the middle of 1962.

**Multiplicative congruence.** The term "congruence" means the following: two integers  $A$  and  $B$  are said to be congruent modulo  $M$ , written  $A \equiv B \pmod{M}$ , if  $A$  and  $B$  have the same remainder when divided by  $M$ . A sequence  $x_1, x_2, \dots$  can be generated by the multiplicative congruence relation, such as  $x_{n+1} \equiv kx_n \pmod{M}$ , as follows: Assume that  $k$  is a given constant and it is multiplied by the first number of the sequence. The result of the multiplication is divided by  $M$  and the remainder of this division process is specified to be the next number of the sequence,  $x_2$ . Thereafter  $x_3$  is obtained by repeating the multiplication and division process using  $x_2$  in place of  $x_1$ . Studies of this process have been made concerning the selection of  $k$  and  $x_1$  relative to  $M$  to ensure a large period. As mentioned by Lehmer, the particular relation given here has the undesirable feature that the right-hand digit positions of the generated numbers are not satisfactory individually as random digits. To overcome this limitation and to try to get a period as large as  $M$ , Rotenberg suggested modifying the relation by adding a constant, that is,  $x_{n+1} \equiv kx_n + c \pmod{M}$ . However, the choice of  $k$  and  $c$  relative to  $M$  not only influences the period but also influences the extent of correlation between successive numbers of the sequence.

**Other recurrence relations.** Other types of recurrence relations, such as using two prior numbers instead of one, have been suggested; and these are reviewed in the paper by Hull and Dobell (1962), as is the method suggested by Rotenberg. However, much more work remains to be done before the behavior of the methods that have been proposed is fully understood. Although pseudo random numbers that pass the statistical tests applied to them have been produced, specific digit positions of these numbers have failed to be satisfactory for use as pseudo random digits. One can, however, get pseudo random digits by judicious use of pseudo random numbers or by putting pseudo random numbers through a compound randomization procedure of the type suggested by Horton and Smith (1949) or Walsh (1949). Studies have been made of the digit positions of irrational numbers such as  $\pi$  and  $e$  as pseudo random digits since they have infinite periods. For example, Metropolis, Reitwiesner, and von Neumann (1950) studied some of the properties of the first 2,000 digits of  $\pi$  and  $e$ , and Stoneham (1965) studied the first 60,000 digits of  $e$ . It is important to keep in mind that although one can pose and apply a large number of meaningful tests

for randomness, any given sequence of numbers or digits cannot be expected to meet all such tests with an arbitrary level of statistical significance.

**Random permutations.** Random permutations represent a special application of random digits in the design of experiments. Here "permutation" is used to indicate that a set of objects has been assigned a specific order. Random permutations are used when it is desired, for example, to assign (order) a set of  $n$  different objects to  $n$  positions on a line at random. The number of different arrangements possible for  $n$  different objects is  $n!$ . For example, three objects labeled 1, 2, 3 can be arranged in  $3!$  ( $3! = 3 \cdot 2 \cdot 1 = 6$ ) different ways: 123, 132, 213, 231, 312, 321.

A process is said to generate random permutations of size  $n$  if each of the  $n!$  different possible permutations is independently generated with equal probability. The probability of observing a specific permutation will be  $1/n!$ . Random permutations of size  $n$  may be based on random digits as follows. If  $n$  is a  $p$ -position number, then  $p$ -position integers will be selected from a table of random digits with the following restrictions: If the number selected is equal to or less than  $n$  and if it has not already been selected for the permutation, then it should be included as part of the permutation; otherwise it should be rejected. Continue this process until all  $n$  places in the permutation have been filled. Although this method is practicable when  $n$  is quite small, there are better methods that do not reject so many random digits (see Cochran & Cox 1950, p. 569; Fisher & Yates 1938, p. 34; Walsh 1957, p. 355).

The idea of having random permutations available as a table is due to George W. Snedecor. Tables of random permutations of size 9 and 16 can be found in Cochran and Cox (1950, p. 577) and Moses and Oakford (1963). For other sources, see Greenwood and Hartley (1962, p. 460).

**Sampling without replacement.** Sampling without replacement is a type of probability sampling of a finite population where each item of the population has an equal and independent probability of being selected but each item is allowed to be included in a sample only once. A random permutation of size  $n$  represents the extreme situation of sampling without replacement where the entire population of size  $n$  is included in the sample in a specified order. If one is interested in selecting, without replacement, random samples of size  $r$  from a population of size  $n$ , then random permutations of size  $n$  can be used to accomplish the selection by using the first  $r$  numbers of each permutation. There are other methods of using ran-

dom numbers to select samples without replacement. For example, a recent set of techniques that can be used on digital computers is due to Fan, Muller, and Rezucha (1962, p. 387). With respect to the direct use of random digits to select samples without replacement, Jones (1959) has studied how many numbers must be used to obtain a sample of given size, taking into account the need to reject duplicate and undesired numbers.

MERVIN E. MULLER

[Other relevant material may be found in ERRORS, article on NONSAMPLING ERRORS, and in the biographies of FISHER, R. A.; and GOSSET.]

#### BIBLIOGRAPHY

- BOX, G. E. P.; and MULLER, MERVIN E. 1958 A Note on the Generation of Random Normal Deviates. *Annals of Mathematical Statistics* 29:610-611.
- BROWN, G. SPENCER 1957 *Probability and Scientific Inference*. London: Longmans.
- BROWN, GEORGE W. 1951 History of RAND's Random Digits: Summary. Pages 31-32 in U.S. National Bureau of Standards, *Applied Mathematical Series, Monte Carlo Method*. Edited by Alston S. Householder et al. Washington: Government Printing Office.
- BROWN, GEORGE W. 1956 Monte Carlo Methods. Pages 279-303 in Edwin F. Beckenbach (editor), *Modern Mathematics for the Engineer*. New York: McGraw-Hill.
- COCHRAN, WILLIAM G. (1953) 1963 *Sampling Techniques*. 2d ed. New York: Wiley.
- COCHRAN, WILLIAM G.; and COX, GERTRUDE M. (1950) 1957 *Experimental Designs*. 2d ed. New York: Wiley.
- DARBISHIRE, A. D. 1907 Some Tables for Illustrating Statistical Correlation. Manchester Literary and Philosophical Society, *Memoirs and Proceedings* 51, no. 16.
- FAN, C. T.; MULLER, M. E.; and REZUCHA, I. 1962 Development of Sampling Plans by Using Sequential (Item by Item) Selection Techniques and Digital Computers. *Journal of the American Statistical Association* 57:387-402.
- FISHER, RONALD A.; and YATES, FRANK (1938) 1963 *Statistical Tables for Biological, Agricultural, and Medical Research*. 6th ed., rev. & enl. New York: Hafner. → Previous editions were published by Oliver and Boyd.
- FLORIDA, UNIVERSITY OF, GAINESVILLE, STATISTICAL LABORATORY 1956 *Symposium on Monte Carlo Methods, Held at the University of Florida . . . March 16 and 17, 1954*. Edited by Herbert A. Meyer. New York: Wiley.
- GOOD, I. J. 1953 The Serial Test for Sampling Numbers and Other Tests for Randomness. Cambridge Philosophical Society, *Proceedings* 49:276-284.
- [GOSSET, WILLIAM S.] (1907-1937) 1943 "Student's" *Collected Papers*. Edited by E. S. Pearson and John Wishart. London: Biometrika Office, University College. → William S. Gosset wrote under the pseudonym "Student."
- GREENWOOD, JOSEPH A.; and HARTLEY, H. O. 1962 *Guide to Tables in Mathematical Statistics*. Princeton Univ. Press. → See especially pages 454-468, "Tables of Random Samples."
- HAMMERSLEY, JOHN M.; and HANDSCOMB, DAVID C. 1964 *Monte Carlo Methods*. New York: Wiley.
- HORTON, H. BURKE; and SMITH, R. TYNES 1949 A Direct Method for Producing Random Digits in Any Number System. *Annals of Mathematical Statistics* 20:82-90.
- HULL, T. E.; and DOBELL, A. B. 1962 Random Number Generators. *SIAM Review* 4:230-254.
- JONES, HOWARD L. 1959 How Many of a Group of Random Numbers Will Be Usable in Selecting a Particular Sample? *Journal of the American Statistical Association* 54:102-122.
- KENDALL, M. G.; and SMITH, B. BABINGTON 1938 Randomness and Random Sampling Numbers. *Journal of the Royal Statistical Society Series A* 101:147-166.
- KENDALL, M. G.; and SMITH, B. BABINGTON 1939 *Tables of Random Sampling Numbers*. Cambridge Univ. Press.
- LEHMER, D. H. 1951 Mathematical Methods in Large-scale Computing Units. Pages 141-146 in Symposium on Large-scale Digital Calculating Machinery, Second, 1949, *Proceedings*. Cambridge, Mass.: Harvard Univ. Press.
- METROPOLIS, N. C.; REITWIESNER, G.; and VON NEUMANN, J. 1950 Statistical Treatment of Values of First 2,000 Decimal Digits of  $e$  and  $\pi$  Calculated on the ENIAC. *Mathematical Tables and Other Aids to Computation* 4:109-111.
- MOOD, ALEXANDER M. 1950 *Introduction to the Theory of Statistics*. New York: McGraw-Hill.
- MOSES, LINCOLN E.; and OAKFORD, ROBERT V. 1963 *Tables of Random Permutations*. Stanford (Calif.) Univ. Press.
- MULLER, MERVIN E. 1958 An Inverse Method for the Generation of Random Normal Deviates on Large-scale Computers. *Mathematical Tables and Other Aids to Computation* 12:167-174.
- RAND CORPORATION 1955 *A Million Random Digits With 100,000 Normal Deviates*. Glencoe, Ill.: Free Press.
- SHREIDER, IULIA A. (editor) (1962) 1964 *Method of Statistical Testing: Monte Carlo Method*. Amsterdam and New York: Elsevier. → First published in Russian.
- STONEHAM, R. G. 1965 A Study of 60,000 Digits of Transcendental  $e$ . *American Mathematical Monthly* 72:483-500.
- TEICHHROEW, DANIEL 1965 A History of Distribution Sampling Prior to the Era of the Computer and Its Relevance to Simulation. *Journal of the American Statistical Association* 60:27-49.
- TIPPETT, LEONARD H. C. (1927) 1952 *Random Sampling Numbers*. Cambridge Univ. Press.
- VON NEUMANN, JOHN 1951 Various Techniques Used in Connection With Random Digits. Pages 36-38 in U.S. National Bureau of Standards, *Applied Mathematics Series, Monte Carlo Method*. Edited by Alston S. Householder. Washington: Government Printing Office.
- WALSH, JOHN E. 1949 Concerning Compound Randomization in the Binary System. *Annals of Mathematical Statistics* 20:580-589.
- WALSH, JOHN E. 1957 An Experimental Method for Obtaining Random Digits and Permutations. *Sankhyā: The Indian Journal of Statistics* 17:355-360.
- WELDON, W. F. R. 1906 Inheritance in Animals and Plants. Pages 81-109 in T. B. Strong (editor), *Lectures on the Method of Science*. Oxford: Clarendon.
- YOUNG, FREDERICK H. 1962 *Random Numbers, Mathematical Induction, Geometric Numbers*. Boston: Ginn.



## RANDOM WALK

See MARKOV CHAINS; QUEUES.

## RANDOMNESS

See EXPERIMENTAL DESIGN; PROBABILITY; RANDOM NUMBERS; SAMPLE SURVEYS; STATISTICS.

## RANK, OTTO

Otto Rank (1884–1939), psychoanalyst and social philosopher, was born of middle-class Jewish parentage in Vienna, lived for many years in Paris, and died in New York a U.S. citizen. At 21 Rank became Freud's brilliant protégé and soon one of the leading psychoanalytic figures. Eventually, he set forth his heretical "birth-trauma" theory; and in 1926, after much soul searching, he finally broke with Freud and left Vienna. Unfortunately, today it is commonly believed, even within the profession, that this theory is Rank's main or sole important contribution. Actually, it plays no role in his far more significant works on art and genetic psychology, which, however, were never to emerge from relative obscurity. A precociously intellectual child in an unhappy family, Rank discovered in adolescence the solace of the theater and later of the other arts. He was then greatly stimulated by Nietzsche and Schopenhauer. Although financial need propelled him toward an engineering career, Rank diverted enough time from his technical studies to begin a book on the psychology of the artist—for long his master theme. He soon came under the influence of Freud's writings and attended lectures by Alfred Adler, through whom, in 1906, Rank was given entree into the little psychoanalytic circle of Vienna. He was at once made its secretary and Freud's personal aide and, efficiently fulfilling both duties, remained at Freud's side for nearly twenty years.

A filial-paternal relationship evolved between "little Rank" and the chief, who gave Rank some financial aid and encouraged him to enter the University of Vienna, from which he received a doctorate in 1912. Benefiting from Freud's criticisms, in 1907 Rank published *Der Künstler* ("The Artist"). There followed a series of psychoanalytic interpretations of myth and literature. Rank made his reputation with *The Myth of the Birth of the Hero* in 1909 and *Das Inzest-Motiv in Dichtung und Sage* ("The Incest-motif in Poetry and Myth") in 1912. His increasingly encyclopedic command of what are now called the "humanities" influenced and supplemented the primarily medical outlook of early psychoanalysis. In 1912 he became coedi-

tor of the new psychoanalytic journals *Imago* and *Internationale Zeitschrift für Psychoanalyse*; and he was naturally included in the secret "praetorian guard" (Rank, Ferenczi, Jones, Sachs, Abraham; later Eitingon) that Freud, worried by the defections of Adler and Jung, had mustered to watch over "the cause."

During World War I Rank edited the *Krakauer Zeitung*, which served as the official journal of the Austrian Army. In Cracow, he met his first wife, Beata Tola Mincer (later a prominent child therapist); they were married in Vienna in 1918.

After the war Rank renewed with even greater vigor his labors for Freud and "the cause." He helped to organize its new publishing house, the Internationaler Psychoanalytischer Verlag, and as its director kept it going under the extremely adverse postwar conditions. He also began for the first time himself to practice psychotherapy, and his interest turned to basic theoretical problems. Between 1920 and 1923, he and Ferenczi had a close intellectual alliance, and together they wrote in 1922 *The Development of Psychoanalysis* (seen by Freud in manuscript, but not published until 1924). It departed somewhat, although not yet fatally, from the Freudian canon.

Within the psychoanalytic movement there was now a certain anomie. The glow of its heroic age was fading; Freud had cancer and was nearing 70. Rivalries for the succession inevitably emerged, and the leading contender may have been Rank. A tense situation already existed in the leadership when Rank published his notorious *Trauma of Birth* (1924a), which was written a little later than the book with Ferenczi, independently and a bit secretly.

It provoked a theoretical scandal which eventually led to a new schism. Rank's essential thesis was that *all* anxiety—and, as a consequence, all neurosis or disposition to neurosis—is a derivation from the original infantile fright upon emerging from the mother. This universal experience does not necessarily result in neurosis but is always more or less traumatic—and is "the ultimate biological basis of the psychical" (Rank [1924a] 1952, p. xlii). All cultural phenomena may be interpreted as either direct expressions of, or strivings to overcome, this primal birth anxiety.

Initially greatly impressed, and describing Rank's idea as "the most important progress since the discovery of psychoanalysis" (Jones 1953–1957, vol. 3, p. 59), Freud gradually cooled when the implications in Rank's theory that were inimical to the hitherto assumed primacy of sexuality and the Oedipus or castration complexes began to be recognized. With Abraham in the lead, the orthodox

party moved to restore theoretical order, Ferenczi before long breaking with Rank. For many months Freud himself tried to patch up the quarrel. But Rank, only half-consciously it seems, had taken a long step toward independence, also evident in his abrupt and successful exploratory sortie to New York that year. Yet the mutual affection and respect between Freud and Rank were still very strong, and it was only after much inner conflict and suffering, and one dramatic but abortive reconciliation, that Rank abandoned his long and distinguished career in Vienna.

In 1926 he established a new therapeutic practice in Paris, where most of his subsequent works were written and where he made the acquaintance of a number of artists and writers, including the novelists Anaïs Nin and Henry Miller.

Rank made many lecturing and teaching journeys to America, settling there permanently late in 1934. For many years, through the support of friends in the social work profession, Rank's psychotherapeutic concepts and methods had been quite influential in America, particularly at the School for Social Work at the University of Pennsylvania. But this was largely fortuitous—Rank's ideas, especially in the nonpsychotherapeutic works after 1929, relate to social work only tangentially.

In 1939 Rank and his wife arranged an amicable divorce, and shortly thereafter he married his secretary, Estelle Buel. A month later Freud died in London, and the following month Rank himself died suddenly of some obscure infection.

**Early studies of myth and literature.** Besides his work on the birth of the hero and on the theme of incest, Rank wrote important treatises on the reappearance in modern literature of the primitive belief in the "twin" or "Double" (1925) and on the "Don Juan" character, as portrayed in Mozart's work and elsewhere (1924b). Although Rank's distinctive later theoretical tendencies can be recognized here and there in what might be called repressed form, his contributions of this period are dominated by the Freudian theory. Its value for the interpretation of all aspects of culture was argued by Rank (with H. Sachs) in *The Significance of Psychoanalysis for the Mental Sciences* (1913).

**Basic theory and technique.** For several years Rank tried to reform the Freudian theoretical and therapeutic apparatus, toward which his attitude grew increasingly critical, in the light of the birth-trauma hypothesis. His major works of this time, written simultaneously, were *Technik der Psychoanalyse* ("The Technique of Psychoanalysis"; 1926–1931) and *Grundzüge einer genetischen Psychologie* . . . ("Outlines of a Genetic Psychology"; 1927–

1929). Midway through these, Rank radically changed his viewpoint.

In the parts written while Rank was still under the influence of the Freudian biocentricism, he insisted on the primacy of the birth trauma, orality, and the relationship (of both boy and girl) to the mother, over Oedipus, or castration, complexes, genitality, and the relationship to the father. Rank declared that neuroses cannot be healed by merely theoretical or historical reconstruction of the past or of the cause of the neurosis; the original (birth) trauma has to be brought into the *present* and actively lived out emotionally. The patient projects upon the analyst the role of the mother and when leaving the analyst is able ideally to experience a second, less traumatic "birth" or psychological separation. Rank also favored short-term analyses and the employment of time limits, stipulated at some point in the process of therapy, for he had begun to believe that too much analytic interpretation might even have a contratherapeutic effect.

It was not until 1928–1929 that Rank drastically changed his conception of the ego, essentially a passive factor in Freud's system. Rank redubbed the ego as the "will"—in actual life experience, he declared, the *primary* psychological reality, a "positive guiding organization and integration of self which utilizes creatively, as well as inhibits and controls the instinctual drives" (Rank [1929–1931] 1945, p. 112). Neither the external (natural or cultural) environment nor the inner biological drives but the active human will is the ultimate causal agency within the individual personality, as well as within all sociocultural phenomena. Moreover, although Rank did acknowledge the existence of strife between culture or repression and the instincts, which Freud had revealed and emphasized, he did not consider it to be of major, continuing, or permanent consequence. Further developing an earlier concept, Rank suggested that by a self-generated (not culturally imposed) "denial" the ego or will can effectively nullify or "annihilate" (that is, not just repress) the instinctual drives, which then cease to play any important role. The true conflict, and ultimate cause of neurosis, lies in a failure of the will, through an excessive sense of guilt.

The normal personality expresses his will in conformance with that of the group; the creative type (often cited by Rank as "the artist"), independently or in opposition; the neurotic, whom Rank regarded as essentially a "failed artist," not at all or at least insufficiently.

In health, art, or neurosis the subjective salvation of the individual's will, in whatever terms it finds expression, is psychologically crucial. Even if



the personality is neurotic, its individuality or autonomy should be respected at all costs. The therapist has to assist the patient in strengthening his will or creative powers. This should culminate in that successful psychological "separation" of patient from therapist, which, as Rank then realized, he had formerly incorrectly interpreted as neonatalist.

In the later volumes of the *Technik* and the *Grundzüge*, Rank abandoned all modes of biological determinism and with his new conception of the will explored a number of conflictful psychological dualities in man, among them "life-fear and death-fear," "total ego and partial ego," "masculine and feminine," "individual and collective." His treatment of this material is at times determinedly "will-ful" in the constructive or optimistic sense; but at others, where the theme of guilt or mortality predominates, dark and subtle, quite in the vein of his old philosophical mentors.

These works have been appropriately compared to Kierkegaard's. For, although the "will" can be, and was by some, interpreted entirely rationalistically, almost from the beginning Rank perceived in the depths of the will a fountain of grace or faith—that mysterious (or "irrational," as we now say) psychological entity, the "soul." Eventually, Rank was able to conceptualize or define the soul as the belief in, or "will to," immortality. Although objectively illusory, this belief was subjectively or psychologically the primary human truth. In his later "psychology of the soul" Rank's notions of "will" and "soul" often overlap and in some contexts are employed interchangeably.

**Art and the artist.** Rank concluded his long investigation of art with *Art and Artist* (1932a)—perhaps his magnum opus. His central theme was that although art, like religion, is to be understood as an expression of man's will to immortality, in art this will is retained in its original individualistic, presocial, prereligious "narcissistic" form. Religion, on the other hand, which has made society possible, is founded on the surrender of the narcissistic mode of consciousness. In spite of this surrender, individuals still long for their original total spiritual freedom, which—by social convention—is periodically revived and celebrated in art. The heroes of art, and indeed often the artists themselves, through trying to retain or regain the primal spiritual omnipotence, may live splendidly but in the end perish tragically.

In form and content, art also reflects the contemporary mode of social faith in immortality (see below), but within this historically varying context the artist always strives to actualize his personal (and the archaic) narcissistic immortality.

As this is in principle antisocial, even though allowed or rewarded by convention, psychological and ideological discords arise between the community and the artist. He inevitably suffers guilt and must overcome severe inner conflicts if he is to create. These conflicts are explored by Rank with great sensitivity and brilliance.

The artist's plight is, however, only an acute or specialized form of a general human conflict, reflected in the existence of art itself and apparently worsening under the pressure of the increasingly antinarcissistic rational ideologies. The convention which had so long protected art seemed to be breaking down for both artist and society. Perhaps, suggested Rank, art itself would have to be sacrificed for some new mode of creativity.

Until this point, Rank had in practice hardly acknowledged the existence of any other mode of creativity but the artistic. He then shifted toward a more general psychology of creativity, which, however, never became entirely "separated" from its former "identification" with art.

**General psychology of culture.** Although devised concurrently with portions of Rank's final work on art, *Psychology and the Soul* was published earlier, in 1930. From the evidence of an originally widespread belief in the "Double" or twin, and from related phenomena, Rank inferred that the original concept of immortality must have been the belief of each individual primitive in his physical and spiritual survival in the form of his own externalized "Double." After this "narcissistic" or "emanistic" era, individual men gradually learned to acknowledge immortalities other than their own. This was the psychological foundation of the next or "animistic era," when immortality was no longer primarily conferred or guaranteed by the personal Double, but rather by the collective spirits of the dead, who animated each embryo spiritually.

The objective facts of procreation were still unknown, or else were resisted. Both narcissism and animism accepted sexuality itself but not the conception of personal or biological procreation, which was then felt to be psychologically destructive. Only later, during the "sexual era," was the psychologically anterior and more natural belief in the immortality of the individual body-soul "Double" abandoned and replaced by a belief in immortality through physical procreation. This was, however, no more than an attempt to conceptualize in biological terms what had always been and remained essentially spiritual.

While hitherto the woman had been only the passive host of the souls of the dead who animated the child, she gradually became identified with the soul force and thereby took her place as its first

biological and first human avatar. This was the psychological basis of the "matriarchal" period of the sexual era, which was followed by a "patriarchal" one (Rank's analysis is here in general accordance with the Bachofenist reconstruction [see BACHOFEN]). Thenceforth, in social ideology the individual father assumed the spiritually and biologically procreative role, and social institutions were remodeled conformably. This meant, however, that each father had to accept what was psychologically unnatural and difficult, namely, that he survived spiritually only vicariously in his child, particularly his son, and thus no longer as his own narcissistic Double entity. After many complex ideological conflicts with the father, according to Rank, the son in turn acquired more and more spiritual importance. The transition from Judaism to Christianity ushered in the spiritual "rule of the child" (Rank [1930] 1950, p. 29), at first in religious and later in modern rationalistic form. The child thus became the final biological symbol of human immortality.

Rank also believed that the modern rational or "psychological era" was analyzing out of existence all forms of spiritual illusion, and that this tendency was epitomized by Freudian thought.

Throughout all historical eras, individuals of both sexes experience profound conflicts both over their changing social ideological roles and over their natural desire to retain that original narcissistic immortality belief that has had to be sacrificed to all forms of social ideology, religious or rational.

Rank then rather complicated this psychohistorical scheme by introducing, in a later chapter, a parallel to the immortality-soul—the will-soul, or "will-god." By this he seems to have meant the soul in the aspect of the belief in temporal omnipotence. The cultural material is now partially retraced in terms of a duality or even conflict between "will" and "soul," or the "two souls." It appears that the survival of the immortality principle required that the temporal will eventually had to be broken and adjudged evil, particularly in its sexual expressions.

The contemporary will-struggle or soul-struggle of the individual, as child or adult, with the community was further explored by Rank in a series of essays, published as *Modern Education* (1932b).

Rank's last work, *Beyond Psychology* (1941), the first to be written in English, was almost complete when he died. His earlier outline of psychohistory, slightly modified here and there, was now reiterated more fully and concretely. The contemporary social malady was examined. Rank believed totalitarianism to be essentially a desperate reac-

tion against the prolonged frustration of man's "irrational" will to immortality. Social catastrophe could be avoided only by allowing this psychological need some better communal means of expression. This seems to be Rank's final counsel.

**Assessment.** In the past decade references to and republications of Rank's works have markedly increased. In 1965 the Otto Rank Association—made up primarily of the social workers he influenced but also including professional people interested in Rank from other points of view—was formed.

It is clear that Rank has enriched the fund of psychotherapeutic theories and techniques. Some of his ideas have been partially absorbed into other repertoires—notably within the field of social-work counseling and by the Rogerian "client-centered," the Perlsian neo-gestalt, and one or two "existentialist" psychotherapies. At present, however, there exists no specifically Rankian school, possibly because of the difficulty of fully orienting either therapist or patient toward Rank's demanding and problematic criterion of achieving artistic (or other) creativity. And in his last works, Rank himself shifted the ultimate arena of decision from the individual—whether normal, artistic, or neurotic—"beyond psychology" to the irrational depths of the collective ideology.

As a psychologist of art and the artist, Rank has few if any peers. On this count, he has drawn tributes even from some of those otherwise indifferent or inimical to his work. Its lasting value in this aspect (although also little appreciated today) seems least disputable.

Perhaps Rank's philosophy or psychology of history is most fruitfully compared with Hegel's. In both, a *Geist* presides over human nature and, therefore, over the changing forms of culture. But Hegel thought that the progressive advance of rationality improved the quality (or "freedom") of the human soul, whereas Rank questioned this.

The error of dismissing the soul or *Geist* as only a "mystification" (fatefully committed by the Young Hegelians, including Marx) has perhaps by now become sufficiently obvious. The necessity of delving beneath the formal socioeconomic phenomena into the irrational psychological realm has meanwhile been stressed by many great sociologists, including Max Weber, Pareto, and Mannheim. Many aspects of their work conceivably may be supplemented with Rank's depth psychology.

Like all systems, Rank's presents a number of unresolved theoretical difficulties, and the following are among those that will have to be coped with. First, supposing that we grant the therapeutic and the experiential value of Rank's more positive



estimate of the role of the human ego or will, yet must or should we abandon the Freudian concept of functional repression and much persuasive evidence that culture is generated and maintained only at the price of *continuing* (and even *increasing*) functional losses in man? Perhaps, however, the object of repression is not solely or primarily biological—psychological qualities may also be inconsistent with culture. Second, Rank often equated artistic and nonartistic modes of creativity and left unclear how we are to differentiate between them. The nature of nonartistic creativity is important, since the greater and more basic part of culture is not a product of art. Third, is the human soul definable *solely* as the belief in immortality? May not the belief in *temporal* omnipotence, and the problems thereof, play an equal or even greater role? Rank's secondary "will-soul" concept may be a partial rendition of this factor. Fourth, Rank's *laissez-faire* or relativistic philosophy generally cannot tell us why any social system or individual way of life should be preferred over another, since in terms of attained soul-belief, if that is present, all are equally true and valuable subjectively and equally illusory objectively. What, if there can or should be any, is the criterion by which we may surmount this relativism? If any Rankian criterion can be inferred, this might plausibly be the degree of the preservation, in some form, of man's earlier and more natural narcissistic soul-beliefs. Then a fifth problem is whether this objective is consistent with an increasingly rationalized civilization—if so, how; and if not, what then?

Should the view gain ground in social thought that, first, the ultimate basis of cultural phenomena is psychological; and, second, that human psychology is not wholly or even primarily explicable in terms of biologically or rationally hedonistic motives, among other consequences a re-examination of Otto Rank's later work should follow. Most current estimates of his place in intellectual history may then have to be revised.

JACK JONES

[For the historical context of Rank's work, see *PSYCHOANALYSIS*, article on *CLASSICAL THEORY*; and the biographies of ADLER; FERENCZI; FREUD. For discussion of the subsequent development of his ideas, see *ANXIETY*; *PSYCHOANALYSIS*, article on *EGO PSYCHOLOGY*.]

#### WORKS BY RANK

- 1907 *Der Künstler: Ansätze zu einer Sexualpsychologie*. Vienna and Leipzig: Heller.  
(1909) 1952 *The Myth of the Birth of the Hero: The Psychological Interpretation of Mythology*. New York:

- Brunner. → First published in German. A paperback edition was published in 1959 by Vintage Books.  
1912 *Das Inzest-Motiv in Dichtung und Sage: Grundzüge einer Psychologie des dichterischen Schaffens*. 2d ed. Leipzig and Vienna: Deuticke.  
(1913) 1916 RANK, OTTO; and SACHS, HANNS *The Significance of Psychoanalysis for the Mental Sciences*. New York: Nervous and Mental Disease Pub. → First published in German. Reprinted in 1964 in the Spring-Summer issue of *Imago* as the "credo" of the new editorial staff.  
(1924a) 1952 *The Trauma of Birth*. New York: Brunner. → First published in German.  
1924b *Die Don Juan-Gestalt*. Leipzig, Vienna, Zurich: Internationaler Psychoanalytischer Verlag.  
(1924) 1925 RANK, OTTO; and FERENCZI, SÁNDOR *The Development of Psychoanalysis*. New York: Nervous and Mental Disease Pub. → First published in German. A paperback edition was published in 1956 by Dover.  
1925 *Der Doppelgänger: Eine psychoanalytische Studie*. Leipzig, Vienna, and Zurich: Internationaler Psychoanalytischer Verlag.  
1926–1931 *Technik der Psychoanalyse*. 3 vols. Leipzig and Vienna: Deuticke. → Volume 1: *Die analytische Situation illustriert an der Traumdeutungstechnik*, 1926. Volume 2: *Gestaltung und Ausdruck der persönlichen Elementen*, 1929. Volume 3: *Die Analyse des Analytikers und seiner Rolle in der Gesamtsituation*, 1931.  
1927–1929 *Grundzüge einer genetischen Psychologie auf Grund der Psychoanalyse der Ich-Struktur*. 3 vols. Leipzig and Vienna: Deuticke. → Volume 1 has no separate title, 1927. Volume 2: *Gestaltung und Ausdruck der Persönlichkeit*, 1928. Volume 3: *Wahrheit und Wirklichkeit: Entwurf einer Philosophie des Seelischen*, 1929.  
(1929–1931) 1945 *Will Therapy and Truth and Reality*. New York: Knopf. → Two books, first published as Volumes 2 and 3 of Otto Rank's *Technik der Psychoanalyse* and Volume 3 of his *Grundzüge einer genetischen Psychologie*. This edition also contains a summary of Volume 1 of *Technik der Psychoanalyse*.  
(1930) 1950 *Psychology and the Soul*. Philadelphia: Univ. of Pennsylvania Press. → First published in German. A paperback edition was published in 1961 by Perpetua Books.  
1932a *Art and Artist: Creative Urge and Personality Development*. New York: Knopf.  
1932b *Modern Education: A Critique of Its Fundamental Ideas*. New York: Knopf.  
(1941) 1958 *Beyond Psychology*. New York: Dover. → The 1941 edition was privately printed in England.

#### SUPPLEMENTARY BIBLIOGRAPHY

- GRINSTEIN, ALEXANDER 1958 *The Index of Psychoanalytic Writings*. Volume 3. New York: International Universities Press. → See pages 1587–1595 for a comprehensive bibliography of Otto Rank's writings.  
JONES, ERNEST 1953–1957 *The Life and Work of Sigmund Freud*. 3 vols. New York: Basic Books. → Volume 1: *Formative Years and the Great Discoveries*, 1953. Volume 2: *Years of Maturity*, 1955. Volume 3: *Last Phase*, 1957. Hostile interpretation of Rank's defection, but indispensable for the earlier period. To be compared with Taft 1958.  
KARPF, FAY B. 1953 *The Psychology and Psychotherapy of Otto Rank*. New York: Philosophical Library.  
NIN, ANAÏS 1966 *The Diary of Anaïs Nin, 1931–1934*. New York: Harcourt; Swallow Press. → Primarily of biographical and literary interest.

PROGOTT, IRA 1956 *The Death and Rebirth of Psychology: An Interpretative Evaluation of Freud, Adler, Jung, and Rank and the Impact of Their Culminating Insights on Modern Man*. New York: Julian Press. → See especially Chapter 7.

TAFT, JULIA JESSIE 1958 *Otto Rank: A Biographical Study*. New York: Julian Press.

VIENNA PSYCHOANALYTIC SOCIETY (1906-1908) 1962 *Minutes of the Society, Volume 1: 1906-1908*. Edited by Herman Nunberg and Ernst Federn. New York: International Universities Press. → Meetings recorded by Rank himself. Note that the editors are now distinctly more respectful toward Rank than has been the Freudian wont.

## RANK CORRELATION

See NONPARAMETRIC STATISTICS, *article on RANKING METHODS*; STATISTICS, DESCRIPTIVE, *article on ASSOCIATION*.

## RANK-SIZE RELATIONS

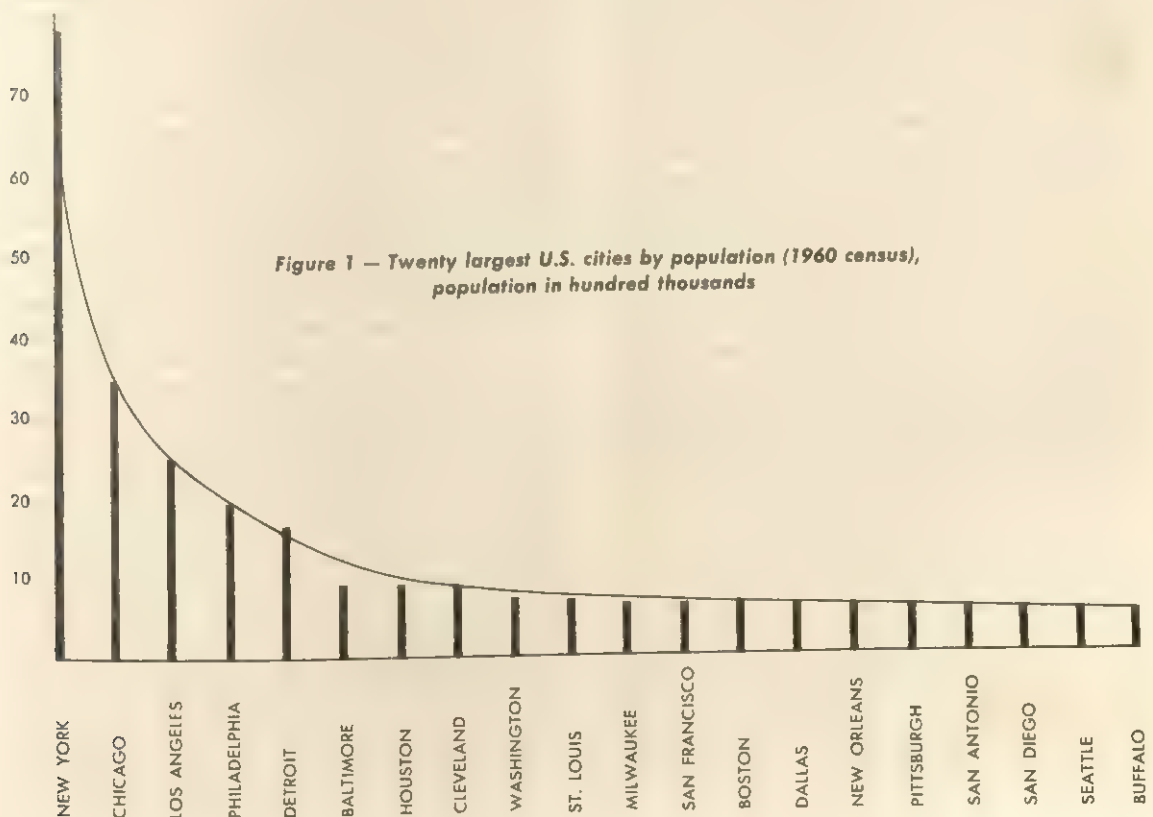
If each object in some collection is characterized by its size, it is possible to rank-order the objects from the largest to the smallest. One may then view the rank as the horizontal axis of a coordinate system on which the objects are arranged,

so that the largest is assigned the horizontal coordinate  $x = 1$ , the next largest  $x = 2$ , etc., assuming no ties. The vertical coordinate,  $y$ , can be viewed as the size of each object. For example, if the objects are cities of the United States rank-ordered according to population (as reported in U.S. Bureau of the Census 1964), New York will have coordinates  $x = 1, y = 7,781,984$ , Chicago will have coordinates  $x = 2, y = 3,550,404$ , San Francisco will have  $x = 12, y = 740,316$ , etc. Such data can be represented on a bar graph. If the number of objects is very large, the bar graph can be approximated by a continuous curve through the pairs of coordinates so defined, as in Figure 1. When ties occur the tied observations are given consecutive ranks.

**Zipf's law.** A question investigated at great length by George K. Zipf (1949) concerned the mathematical properties of rank-size curves obtained from collections of many different sorts of objects. In particular, Zipf examined the rank-size curves of cities (by population), biological genera (by number of species), books (by number of pages), and many other collections.

Zipf found that in most instances the rank-size

POPULATION





curves were very nearly segments of rectangular hyperbolas, that is, curves whose equations are of the form  $xy = \text{constant}$ , or, as expressed in logarithmic coordinates,  $\log x + \log y = \text{constant}$ . Therefore, when such rank-size curves are plotted on log-log paper they are very nearly straight lines with slopes close to  $-1$ . At least, this was the case with the collections that Zipf singled out for attention. The relation  $\text{rank} \times \text{size} = \text{constant}$  is sometimes referred to as Zipf's law.

*The principle of least effort.* Zipf attempted to derive his law from theoretical considerations, which he summarized in the so-called principle of least effort. The connection between this principle and the rank-size law is by no means clear, and Zipf's theoretical arguments now have at most only historical interest. However, his work attracted wide attention and spurred investigations more rigorous and theoretically more suggestive than his own.

The following discussion will concern Zipf's law in its statistical-linguistic context. Let the collection be the words used in some large verbal output, say in a book or in a number of issues of a newspaper. Such a collection is called a "corpus." Let the size assigned to each word be the number of times the word appears in the corpus. Then in most cases the "largest" (that is, the most frequently occurring) word will be *the*, the next "largest" will be *and*, etc. (Thorndike & Lorge 1944). By the nature of the ordering, the curve will be monotonically decreasing (a J curve)—at first steeply falling, then gradually flattening out—for by the time the low frequency words are reached, there will be many having the same (small) number of occurrences. Their ranks, however, will keep increasing, because in a rank-size graph objects having the same size are nevertheless assigned consecutive ranks. Therefore, larger and larger blocks of the bar graph will have the same height, which means that the continuous curve through the bar graph will become increasingly flatter. The rectangular hyperbola, being asymptotic to the horizontal axis, also has this property, which in part accounts for the good agreement between the hyperbola and the rank-size graph.

It is known that the most frequently occurring words in any language are usually the shorter ones. In English, for example, these are articles, prepositions, and conjunctions. If the use of a word represents effort by the speaker, and if the speaker tries to minimize effort, he can be expected to use the shortest words most frequently. But there are comparatively few short words, because there are comparatively few combinations

to be formed from few letters (or phonemes). Consequently, the number of different high-ranking words (those with low rank numbers) will be small. By a similar argument, the words well along in the ranking are, on the average, the longer ones, and so their numbers will be large; thus, many will have equal frequencies. It is, in fact, observed that the largest number of different words in a typical corpus of some thousands of words are those used only once.

In developing his argument from the principle of least effort in the context of language statistics, Zipf postulated opposing tendencies on the part of the speaker and on the part of the hearer. From the point of view of the speaker, so his argument goes, the language most economical of effort would be one with very few words, each word having many meanings. From the point of view of the hearer, on the contrary, the ideal language would be one in which each word has a unique meaning, since in that case the labor of matching meaning with context would be saved. A balance is struck between the effort-economizing tendency of the speaker and that of the hearer by a certain distribution of *ranges* of meaning associated with the distribution of frequencies with which words are used.

Whatever merit Zipf's principle of least effort has in the context of language statistics, his arguments to the effect that the same principle is responsible for the rank-size distributions of the great variety of collections from cities (ranked by populations) to applicants for marriage licenses (ranked by distances between the homes of the bride and the groom) seem of questionable relevance. It is not clear how the principle of least effort operates in each instance to produce the observed rank-size curve. Similar distributions may be traceable to mathematically isomorphic processes, but Zipf, in his principle of least effort, emphasizes not the possible mathematical genesis of his law but its alleged origins in the nature of human behavior.

That this can be misleading can be seen in a hypothetical case of a scholar who, having noted that the weights of beans, of rabbits, and of people are normally distributed, concludes that the normal distribution is a manifestation of a "life force" (because beans, rabbits, and people are biological objects) and seeks the manifestation of this "force" in all instances where the normal distribution is observed. The normal distribution may arise, and presumably often does, from a certain kind of interplay of chance events (roughly, addition of many nearly independent and not wildly dissimilar

random variables). Thus, insofar as the normal distribution arises in many contexts, its genesis may be the common statistical structure of the contexts, not their contents—that is, not whether they involve beans or people, animate or inanimate objects. [See PROBABILITY, *article on* FORMAL PROBABILITY, for a discussion of how the normal distribution arises from the interplay of chance events; this is called the central limit theorem.]

### Statistical rationales for Zipf's law

Explanations of Zipf's law and of rank-size relations in general, like the explanations of the normal distribution discussed above, are to be sought in the statistical structure of the events that might generate these relations instead of in the nature of the objects to which the relations apply.

Consider a frequency density in which the horizontal axis represents nonnegative size while the vertical axis represents the relative frequency with which objects of a given size are encountered in some large collection. Call this frequency density  $f(x)$ , so that  $\int_0^\infty Nf(x) dx = N$ , where  $N$  is the number of objects in the collection. Consider now  $G(x) = \int_x^\infty Nf(t) dt$ , which is the number of objects having sizes greater than  $x$ . But this is exactly the rank (according to size) of the object whose size is  $x$ , assuming that the ranks of objects of equal size have been assigned arbitrarily among them. Therefore the rank-size curve is essentially the integral of the size-frequency curve. Any mathematical theory that applies to the one will apply to the other after the transformation just described has been performed. In particular, Zipf's law holds ( $G(x) = K/x$ ) if and only if the frequency density,  $f(x)$ , is of form  $K/x^2$ , since  $\int_x^\infty (K/t^2) dt = K/x$ .

Investigators seeking a statistical rationale for Zipf's rank-size curves studied the associated frequency density curves. Probabilistic models underlying some common size-frequency distributions are well known. For example, if there is a collection of objects characterized by sizes, and the size of each object is considered as the sum of many random (at least approximately independent) variables, none of which has a dispersion dominating the dispersions of others, then the resulting distribution of sizes will be approximately a normal distribution. This probabilistic process adequately accounts for the normal distributions so frequently observed in nature. On the other hand, suppose that each object suffers repeated increments or decrements that are proportional to the sizes of the objects on which they impinge. Then the ultimate equilibrium distribution will be a so-called loga-

rithmic normal one, that is, a frequency distribution of a random variable whose logarithm is normally distributed. Such frequency distributions are also commonly observed [see DISTRIBUTIONS, STATISTICAL, *article on* SPECIAL CONTINUOUS DISTRIBUTIONS].

The problem of finding a statistical rationale for Zipf's law, therefore, is that of finding a probabilistic process that would result in an equilibrium distribution identical with the derivative of Zipf's rank-size curve—that is, one where the frequency would be inversely proportional to the square of the size.

**Models yielding Yule distributions.** Herbert Simon (1955) proposed such a model for the size-frequency distribution of verbal outputs. The essential assumption of the model is that as the corpus is being created (in speaking or writing) the probability of a particular word being added to the already existing list is proportional to the total number of occurrences of words in that frequency class and, moreover, that there is also a nonnegative probability that a new word will be added to the list. In some variants of Simon's model, the latter probability is a constant; in others it decreases as the corpus grows in size (to reflect the depletion of the vocabulary of the speaker or writer). The resulting equilibrium frequency distribution coincides with Zipf's, at least in the high-frequency range.

Moreover, it appears that similar models are plausible rationales for many other distributions. Essentially it is assumed that the increments impinging on objects are proportional to the sizes of the objects and also that new objects are added to the population according to a certain probability law. The first assumption leads to a logarithmic normal distribution. Combined with the second assumption (the "birth process") it leads to so-called Yule distributions, which greatly resemble Zipf's.

If the derivation indicated by Simon is accepted, the principle of least effort becomes entirely superfluous, for clearly it is the probabilistic structure of events rather than their content that explains the frequency density and hence the rank-size distributions that Zipf considered to be *prima facie* evidence for the principle of least effort.

**Information-theoretical models.** The principle of least effort was not entirely abandoned by those who sought rationales for Zipf's law, at least for its manifestation in language statistics. Benoît Mandelbrot (1953), for example, restated the principle of least effort as follows. Assume, as Zipf did, that there is an effort or cost associated with each word. Then if the speaker is to economize effort,



clearly he should select the cheapest word and speak only that word. However, discourse of this sort would not convey any information, since if the same word is spoken on all occasions, the hearers know in advance what is going to be said and get no new information from the message. The problem is, according to Mandelbrot, not to minimize effort (or cost, as he calls it) *unconditionally* but rather to minimize it *under the constraint* that a certain average of information per word must be conveyed. Equivalently, the problem is to maximize the information per word to be conveyed under the constraint that a certain average cost of a word is fixed.

Here Mandelbrot was able to utilize the precise definition of the *amount of information* conveyed by a message, as formulated in the mathematical theory of communication (Shannon & Weaver 1949). Having cast the problem into mathematical form, Mandelbrot was able to derive Zipf's rank-size curve as a consequence of an assumption related to the principle of least effort, namely, the minimization of cost, given the amount of information to be conveyed [see INFORMATION THEORY].

Actually, Mandelbrot's derived formula was more general than Zipf's. While Zipf used rank  $\times$  size = constant, Mandelbrot obtained from his model the formula

$$P_r = P(r + m)^{-B},$$

in which  $P_r$ , being the frequency of occurrence, represents size,  $r$  is rank, and  $P$ ,  $m$ , and  $B$  are constants. If  $B = 1$  and  $m = 0$ , Mandelbrot's formula reduces to Zipf's rank-size law.

As would be expected, the generalized formula fits most rank-size verbal-output curves better than Zipf's; in addition, it is derived rigorously from plausible assumptions.

**Investigations of rank-size relations.** Zipf himself suggested a generalization of his rank-size law, namely,

$$(\text{rank})^q \times \text{size} = \text{constant}, \quad q > 0.$$

With  $q$  as an extra free parameter, clearly more of the observed rank-size curves could be fitted than without it. Of greater importance to a rank-size theory is a rationale for introducing the exponent  $q$ . However, Zipf's arguments on this score are as vague as those related to the originally postulated law.

Frank A. Haight (1966) pointed out the dependence of the observed size-frequency relationship on the way the data are grouped. Suppose, for example, the size-frequency relationship of cities is examined. Clearly, in order to obtain

several cities in each population class the populations must be rounded off—say, to the nearest thousand or ten thousand. Haight has shown that if  $Z$  is the number of digits rounded off in the grouping and if Zipf's generalized rank-size law, given above, holds for cities, then the size-frequency distribution will be given by

$$p_n(Z) = (2Z)^{-1/q} \left\{ \left[ \frac{2Z}{2n-1} \right]^{1/q} - \left[ \frac{2Z}{2n+1} \right]^{1/q} \right\},$$

where  $[x]$  is the integral part of  $x$ . Here  $p_n(Z)$  is the fraction of cities with population near  $n$  (rounded off by  $Z$  digits).

As  $Z$  becomes large, this distribution tends to the zeta distribution, namely,

$$p_n = (2n-1)^{-1/q} - (2n+1)^{-1/q}.$$

The zeta distribution gives a fairly good fit to the populations of the world metropolitan areas rounded off to the nearest million. (There are 141 such areas with populations close to one million, 46 with populations close to two million, etc.) The number of accredited colleges and universities in the United States with student populations rounded off to the nearest thousand is also well fitted by the zeta distribution.

Robert H. MacArthur (1957) treated the problem of the relative abundance of species in a natural organic population by means of a model isomorphic to a random distribution of  $n-1$  points on a line segment. The distances between the points then represent the "sizes of biological niches" available to the several species and therefore the abundance of the species. Thus the size of the  $r$ th rarest species among  $n$  species turns out to be proportional to

$$\sum_{i=1}^r (n-i+1)^{-1}.$$

It appears, then, that the original rank-size law proposed by Zipf is only one of many equally plausible rank-size laws. Clearly, if objects can be arranged according to size, beginning with the largest, *some* monotonically decreasing curve will describe the data. The fact that many of these curves are fairly well approximated by hyperbolas proves nothing, since an infinitely large number of curves resemble hyperbolas sufficiently closely to be identified as hyperbolas. No theoretical conclusion can be drawn from the fact that many  $J$  curves look alike. Theoretical conclusions can be drawn only if a rationale can be proposed that implies that the curves *must* belong to a certain class. The content of the rationales becomes, then, the content-bound theory. Specifically, the constants

contained in the proposed mathematical model can receive a content interpretation.

For example, in Mandelbrot's model 1/B is interpreted as the "temperature" of the verbal output. Taken out of context the "temperature of a language sample" seems like an absurd notion. But the term is understandable in the meaning of the exact mathematical analogue of temperature, as the concept is derived in statistical thermodynamics. In this way, formal structural connections are established between widely different phenomena, which on a priori grounds would hardly have been suspected to be in any way related.

Such discoveries are quite common. A strict mathematical analogy was found between the distribution of the number of bombs falling on districts of equal areas in London during the World War II bombing and the distribution of the numbers of particles emitted per unit time by a radioactive substance. The unifying principle is to be found neither by examining bombs nor by examining radioactive substances but rather by inquiring into the probabilistic structure of the events in question.

Having noted that the rank-size relation is simply another way of viewing the size-frequency relation, it can be seen that all the studies of the latter are relevant to the former. Lewis F. Richardson (1960) gathered extensive data on the incidence of "deadly quarrels"—that is, wars, riots, and other encounters resulting in fatalities [see the *biography* of RICHARDSON]. Designating the size of a deadly quarrel by the logarithm of the number of dead, he studied the associated size-frequency relation, seeking to derive a law of "organization for aggression." He believed he had found evidence for such a law in the circumstance that the size-frequency relation governing Manchurian bandit raids was very similar to the one governing Chicago gangs in the prohibition era. Here one might also interpose the objection that the similarity may have nothing to do with aggression as such, being simply a reflection of the probabilities governing the formation, growth, and dissolution of human groups. Curiously, a comparison of the distribution of sizes of casual groups such as people gathered around swimming pools (Coleman & James 1961) turned out to be different from that of gangs. Moreover, the former type is derived from a stochastic process in which single individuals can join or leave a group, while the latter derive from a process in which no individual can leave unless the whole group disintegrates (Horvath & Foster 1963). These results are intriguing, because the difficulty with which an

individual may leave a gang is well known, and so Richardson's conjecture may have been not without foundation.

It appears, therefore, that the search for the stochastic processes underlying observed rank-size or size-frequency relations can result in important theoretical contributions.

ANATOL RAPOPORT

#### BIBLIOGRAPHY

- COLEMAN, JAMES S.; and JAMES, JOHN 1961 The Equilibrium Size Distribution of Freely-forming Groups. *Sociometry* 24:36-45.
- HAIGHT, FRANK A. 1966 Some Statistical Problems in Connection With Word Association Data. *Journal of Mathematical Psychology* 3:217-233.
- HORVATH, WILLIAM J.; and FOSTER, CAXTON C. 1963 Stochastic Models of War Alliances. *General Systems* 8:77-81.
- MACARTHUR, ROBERT H. 1957 On the Relative Abundance of Bird Species. National Academy of Sciences, *Proceedings* 43:293-295.
- MANDELBROT, BENOÎT 1953 An Informational Theory of the Statistical Structure of Language. Pages 486-502 in Willis Jackson (editor), *Communication Theory*. New York: Academic Press; London: Butterworth. → Contains three pages of discussion of Mandelbrot's article.
- PIELOU, E. C.; and ARNABON, A. NEIL 1966 Correction to One of MacArthur's Species-abundance Formulas. *Science* 151:592 only. → Refers to MacArthur 1957.
- RAPOPORT, ANATOL 1957 Comment: The Stochastic and the "Teleological" Rationales of Certain Distributions and the So-called Principle of Least Effort. *Behavioral Science* 2:147-161.
- RICHARDSON, LEWIS F. 1960 *Statistics of Deadly Quarrels*. Pittsburgh: Boxwood.
- SHANNON, CLAUDE E.; and WEAVER, WARREN (1948-1949) 1959 *The Mathematical Theory of Communication*. Urbana: Univ. of Illinois.
- SIMON, HERBERT A. 1955 On a Class of Skew Distribution Functions. *Biometrika* 42:428-439. → Reprinted in Simon's *Models of Man*.
- THORNDIKE, EDWARD L.; and LORGE, IRVING 1944 *The Teacher's Word Book of 30,000 Words*. New York: Columbia Univ., Teachers College.
- U.S. BUREAU OF THE CENSUS 1964 Population and Land Area of Urbanized Areas: 1960 and 1950. Table 22 in U.S. Bureau of the Census, *Census of Population: 1960. Volume 1: Characteristics of the Population. Part 1*. Washington: Government Printing Office.
- ZIFF, GEORGE K. 1949 *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology*. Reading, Mass.: Addison-Wesley.

#### RANKE, LEOPOLD VON

Leopold von Ranke (1795-1886) is one of the great figures of European nineteenth-century scholarship and a founder of modern historical science. The son of a lawyer in a small town of Thuringia, he graduated from Schulpforta, one of the most



renowned public schools of Germany, and studied philology and theology at the University of Leipzig. In 1818 he became a teacher of classical languages in the high school (*Gymnasium*) in Frankfurt on the Oder. In 1824 his first book, the *History of the Latin and Teutonic Nations*, appeared and immediately brought him wide recognition. Ranke was appointed professor at the University of Berlin and received a travel grant from the Prussian government that permitted him to spend four years abroad, mainly in Italy—decisive years for the development of his historical views. He returned to Berlin in 1831 and, with the exception of extended trips for research in German, French, and English archives, he spent the rest of his life in Berlin. In the historical seminars that he made an essential part of the education of a young historian he trained most of the better-known German historians of the nineteenth century. Ranke retired from teaching in 1871 but continued to work on a last great enterprise, a world history. When he died in 1886, ennobled by the king of Prussia and laden with honors from all countries of the world, he was generally recognized as the greatest historical scholar of the modern world.

Ranke's collected works comprise 54 volumes. Most famous among them are his history of the popes (1834–1836), his history of the Reformation in Germany (1839–1847), his works on French history (1852) and on English history (1859–1869), and his 12 volumes on Prussian history (1847–1848). All of them focus on developments from the fifteenth to the eighteenth century; they show that Ranke's main attainment was to provide a scientific basis for the study of modern history.

Ranke's most important innovation was the introduction of a critical historical method. In an appendix to his first work, the *History of the Latin and Teutonic Nations*, which was concerned with the development of a European state system around 1500, Ranke showed that the historical works by contemporaries, on which previous treatments of this period were based, were vitiated by personal and political prejudices and should be used only with great caution. When Ranke was in Italy he discovered the reports that Venetian ambassadors delivered before the Senate after their return from their diplomatic missions and realized that such materials of an official character, produced in the course of the conduct of affairs, were far superior to narrative sources as tools for discovering the truth about the past. Thus he established that the materials from which the historian should construct his history ought to be, wherever possible, documentary sources. Both the proposition that

serious history ought to be based on archival research and the large-scale publications of documentary source materials that were started in the nineteenth century have their origin in Ranke's adoption of a new critical method.

This methodological innovation sprang from Ranke's general notions about the task of the historian: "to show what actually happened" (*wie es eigentlich gewesen*). This statement sounds simple and matter-of-course, but it was meant to be a challenge to the philosophies of history of the first half of the nineteenth century, particularly the philosophy of Hegel. According to Ranke, history has no final aim that can be abstractly defined and is not an ascending process in which the later period is always superior to the earlier one. Ranke thought that it is almost sacrilegious for man to believe that he can grasp God's providential plan, but he did believe that history helps reveal the working of God by demonstrating the richness and variety of life. Thus the statement that the historian ought "to show what actually happened" was complementary to another famous pronouncement by Ranke—that "each period is equally close to God."

For Ranke each period of history has its own individual features: it is unique. Each period, almost each historical phenomenon, reflects a distinctive "idea." His own deeply religious feeling, his romantic enthusiasm for the "abyss of individuality," as well as Platonic influences—all combined to form Ranke's view of the role of "ideas" in history. He had a very fine understanding of the manner in which the various activities of a period—political, literary, intellectual—are permeated by the same spirit and express the same "idea."

Despite his understanding of the interrelation of all these spheres, Ranke was chiefly a political and diplomatic historian. He belonged to the age of rising nationalism, and his interests were focused on the great powers that were the political embodiments of the spirits of the various European nations. He saw these powers as individualities, as expressions of different "ideas." The clearest formulation of these views can be found in his essay entitled "A Dialogue on Politics," published in 1836 in the periodical *Historisch-Politische Zeitschrift*, which Ranke himself edited and which represented one of his few ventures into the field of practical politics. In the essay Ranke defended the existing governments against the revolutionary movements of 1830 and explained that liberalism could not set a generally valid political pattern because each state was a living organism, an individuality, and must have its own particular institutional forms.

The events of foreign policy must form the central interest of the historian because the great powers developed their particular individualities during, and by means of, struggles against each other. Ranke never used the expression "primacy of foreign policy," but the doctrine that is signified by this term—that external pressure forms and determines the internal structure of a state—is implied in his works. Because Ranke was principally concerned with foreign policy, he had little understanding of the importance of the changes that industrialization brought about in his own century. Ranke was fundamentally conservative.

Ranke's views are incompatible with the aims of the modern social sciences. He rejected the possibility of laws of social development and of patterns generally applicable to social action or behavior. He was a great writer, and his books are not simply histories but also works of literature. Nevertheless, the turn he gave to the development of historical scholarship did have an influence on the development of the social sciences in that his views were a decisive factor in "professionalizing" history. History became an academic subject that required specialized training, and archival research and the editing of source materials became a great part of the activity of a historian. Although originally such efforts focused on sources for the history of foreign affairs, they soon extended to other aspects of the past: the sources for institutional, economic, and social developments. Thus historical scholarship has produced significant material for all kinds of social research. Moreover, by emphasizing the particular and individual character of each period of the past, Ranke implicitly suggested the existence of a relativity of values and helped to remove the barriers that had prevented an understanding of foreign cultures.

FELIX GILBERT

[Directly related are the entries HISTORIOGRAPHY; HISTORY. Other relevant material may be found in the biography of ROBINSON.]

#### BIBLIOGRAPHY

Von Laue 1950 gives a description of Ranke's early intellectual development and contains a bibliographical essay. The intellectual roots of Ranke's views on history have been analyzed in Meinecke 1936. See Vierhaus 1957 for Ranke's views on social developments and social history, and for Ranke's part in the professionalization of history see Higham et al. 1965.

#### WORKS BY RANKE

(1812-1885) 1984 *Aus Werk und Nachlass*. Volume 1: Tagebücher. Edited by Walther Peter Fuchs. Munich: Historische Kommission bei der Bayerischen Akademie der Wissenschaften. → The first volume of a projected set of Ranke's previously unpublished works.

- (1814-1886a) 1949 *Neue Briefe*. Edited by Bernhard Hoesft and Hans Herzfeld. Hamburg (Germany): Hoffmann & Campe.
- (1814-1886b) 1949 *Das Briefwerk*. Edited by Walther P. Fuchs. Hamburg (Germany): Hoffmann & Campe.
- (1824) 1909 *History of the Latin and Teutonic Nations (1494-1514)*. London: Bell. → First published as *Geschichte der romanischen und germanischen Völker von 1494 bis 1514*.
- (1833) 1950 *The Great Powers*. Pages 181-218 in Theodore H. Von Laue, *Leopold Ranke: The Formative Years*. Princeton Univ. Press. → First published as "Die grossen Mächte."
- (1834-1836) 1912 *The History of the Popes During the Last Four Centuries*. 3 vols. London: Bell. → First published as *Die römischen Päpste in den letzten 4 Jahrhunderten*.
- (1836) 1950 *A Dialogue on Politics*. Pages 152-180 in Theodore H. Von Laue, *Leopold Ranke: The Formative Years*. Princeton Univ. Press. → First published as "Politisches Gespräch" in Volume 1/2 of the *Historisch Politische Zeitschrift*.
- (1839-1847) 1905 *History of the Reformation in Germany*. New York: Dutton. → First published as *Deutsche Geschichte im Zeitalter der Reformation*.
- (1847-1848) 1930 *Zwölf Bücher preussischer Geschichte*. 5 vols. Munich: Historische Kommission bei der Bayerischen Akademie der Wissenschaften. → An enlarged edition of the *Neun Bücher preussischer Geschichte*.
- 1852 *Civil Wars and Monarchy in France, in the Sixteenth and Seventeenth Centuries: A History of France Principally During That Period*. 2 vols. London: Bentley. → First published as *Französische Geschichte in five volumes*. The English translation was never completed.
- (1859-1869) 1966 *A History of England Principally in the Seventeenth Century*. 8 vols. Oxford: Clarendon. → First published as *Englische Geschichte, vornehmlich im sechzehnten und siebzehnten Jahrhundert*.
- Sammtliche Werke*. 34 vols. 3d ed. Leipzig: Duncker & Humblot. 1875-1900.
- (1881-1888) 1882-1902 *Weltgeschichte*. 9 vols. Leipzig: Duncker & Humblot.

#### SUPPLEMENTARY BIBLIOGRAPHY

- HIGHAM, JOHN, KRIEGER, LEONARD, and GILBERT, FELIX. 1965 *History*. Englewood Cliffs, N.J.: Prentice-Hall.
- IGGERS, GEORGE G. 1962 *The Image of Ranke in American and German Historical Thought*. *History and Theory* 2: 17-40.
- MEINECKE, FRIEDRICH. (1936) 1950 *Werke*. Volume 3: *Die Entstehung des Historismus*. Munich: Oldenbourg.
- VIERHAUS, RUDOLF. 1957 *Ranke und die soziale Welt*. Münster (Germany): Aschendorff.
- VON LAUE, THEODORE H. 1950 *Leopold Ranke: The Formative Years*. Princeton Univ. Press.

#### RANKING METHODS

See under NONPARAMETRIC STATISTICS.

#### RAPAPORT, DAVID

Although David Rapaport (1911-1960) died at the early age of 49, he left an enduring legacy to the fields of psychoanalysis and psychology. He was born to a middle-class Hungarian family but



in his early teens was already deeply involved in left-wing political activities and spent several years in Israel working as a surveyor in a kibbutz. He returned to Hungary, resumed his studies at the Péter Pázmány University in Budapest, entered psychoanalysis, and ultimately changed his field of study from physics to psychology. His effort to integrate the psychosocial investigations of Erik H. Erikson with psychoanalytic theory undoubtedly derived some of its impetus from his early period of political activity.

Rapaport immigrated to America in 1938 and worked principally in two private psychiatric hospitals—the Menninger Clinic of Topeka, Kansas, from 1940 to 1948, and the Austen Riggs Center at Stockbridge, Massachusetts, from 1948 until his death.

His central interest was the psychology of thinking, both its relatively autonomous laws and the influence upon it of the affective and drive life. In his formulation the central problem in the psychology of thinking is an epistemological one: how man can come to know reality, influenced as he is by his drives.

His first major work was an integration of the psychoanalytic, clinical and experimental psychological, and psychopathological data relating to the concepts of emotions and memory (1942). He proposed a hierarchy of organizing principles of the thought process, shading from drive into ego motivations, and described memory, concept formation, and anticipation as major dimensions of thinking.

This concern with the nature of thinking was continued in his pioneering work in diagnostic psychological testing. In detailed studies of the rationales of specific tests, he showed that while some of these tests are more revealing of intellectual functioning and others more revealing of personality characteristics, all recorded responses have implications for both of these aspects of human functioning, and the two kinds of tests have synergistic value when jointly employed in batteries (Rapaport et al. 1945–1946).

The next major period of Rapaport's life was spent on a creative systematization of psychoanalytic theory. He explicated the basic Freudian model (1951a; Rapaport & Gill 1959) and the Freudian theory of thinking (1951b) and of affect (1953b). However, he was also a systematizer of the newer ego psychology (1959a), to which he believed the contributions of Heinz Hartmann (1939) and Erik H. Erikson (1959) to be the most important. He saw Hartmann's contributions as a programmatic outline for a comprehensive psy-

chology, which would encompass issues of adaptation as well as the more usual psychoanalytic considerations, and Erikson's main import to be his specification of the stages of psychosocial epigenesis, with Erikson's concepts of mode and modality providing a bridge between individual and social psychology [see IDENTITY, PSYCHOSOCIAL; see also Erikson 1950].

Rapaport's more innovative contributions (1958) lay in developing the concept of the autonomy of the ego from the environment and in elaborating the relationship between this autonomy and the autonomy of the ego from the id, a concept previously formulated by Hartmann. He integrated the dimensions of activity-passivity into the theory of ego autonomy (1953a), resuscitated Freud's theory of consciousness as a "sense organ," and developed a theory of states of consciousness (1957).

Rapaport worked toward integrating psychoanalysis with contributions from general psychology, notably the work of Jean Piaget (Wolff 1960), from ethology, and from experimental studies relating innate central organismic states and environmental input (1960a). His systematization of psychoanalytic theory was capped by a monograph which has the merit of attempting to translate much of psychoanalytic thinking—on the "metapsychological" level—into the discourse of contemporary psychology (1959b).

Perhaps even more central than any of these contributions is Rapaport's emphasis and elaboration in all his work of Freud's concept of structure in psychological functioning. Rapaport defined structures as processes of slow change, in contrast to motives, which are processes of rapid change. He believed that the sorely needed psychoanalytic theory of learning would deal with the establishment and maintenance of structures and that this theory could be based on Freud's hypotheses concerning the functioning of attention cathexes. He included cognitive structures among the most critical determinants of behavior.

The last phase of Rapaport's work was an experimental program concerning structure formation. Except for a brief statement included in his study "On the Psychoanalytic Theory of Motivation" (1960b) and a paper on "The Theory of Attention Cathexis," this experimental work is available only in the publications of his students and associates (Paul 1959; Schwartz & Rouse 1961).

Rapaport insisted that in psychoanalytic theory "instinctual drive" is a *psychological* concept that cannot be equated with either the peripheral somatic or the neurophysiological concepts of drives in behavioristic theories. He argued that a self-

contained psychology can and should be developed in its own terms.

Rapaport was a complex and colorful man of brilliant mind, strong opinions, and many interests. Despite stylistic roots in rabbinical casuistry, his condensed writing unfailingly strikes to the heart of an issue. He was an inspiring, dedicated, and demanding teacher, responsible for launching and contributing to the careers of a number of leading workers.

MERTON M. GILL AND GEORGE S. KLEIN

[For discussion of the subsequent development of Rapaport's ideas, see CLINICAL PSYCHOLOGY; IDENTITY, PSYCHOSOCIAL; PSYCHOANALYSIS.]

#### WORKS BY RAPAPORT

- (1942) 1950 *Emotions and Memory*. 2d ed. New York: International Universities Press.
- 1945-1946 RAPAPORT, DAVID; SCHAFFER, R.; and GILL, M. M. *Diagnostic Psychological Testing: The Theory, Statistical Evaluation, and Diagnostic Application of a Battery of Tests*. 2 vols. Chicago: Year Book Publishers.
- 1951a The Conceptual Model of Psychoanalysis. *Journal of Personality* 20:56-81.
- (1951b) 1959 RAPAPORT, DAVID (editor and translator) *Organization and Pathology of Thought: Selected Sources*. Austen Riggs Foundation, Stockbridge, Mass., Monograph No. 1. New York: Columbia Univ. Press.
- 1953a Some Metapsychological Considerations Concerning Activity and Passivity. Unpublished manuscript. → Two lectures given at the staff seminar of the Austen Riggs Center.
- 1953b On the Psycho-analytic Theory of Affects. *International Journal of Psycho-analysis* 34:177-198.
- 1957 Cognitive Structures. Pages 157-200 in Colorado, University of, Psychology Department, *Contemporary Approaches to Cognition: A Symposium*. Cambridge, Mass.: Harvard Univ. Press.
- 1958 The Theory of Ego Autonomy: A Generalization. *Menninger Clinic, Bulletin*, 22:13-35.
- 1959a A Historical Survey of Psychoanalytic Ego Psychology. *Psychological Issues* 1, no. 1:5-17.
- 1959b The Structure of Psychoanalytic Theory: A Systematizing Attempt. Pages 55-183 in Sigmund Koch (editor), *Psychology: A Study of a Science*. Volume 3: Formulations of the Person and the Social Context. New York: McGraw-Hill.
- 1959 RAPAPORT, DAVID; and GILL, MERTON M. The Points of View and Assumptions of Metapsychology. *International Journal of Psycho-analysis* 40:153-162.
- 1960a Psychoanalysis as a Developmental Psychology. Pages 209-255 in Bernard Kaplan and Seymour Wapner (editors), *Perspectives in Psychological Theory: Essays in Honor of Heinz Werner*. New York: International Universities Press.
- 1960b On the Psychoanalytic Theory of Motivation. Volume 8, pages 173-247 in Marshall R. Jones (editor), *Nebraska Symposium on Motivation*. Lincoln: Univ. of Nebraska Press.

#### SUPPLEMENTARY BIBLIOGRAPHY

- ERIKSON, ERIK H. (1950) 1964 *Childhood and Society*. New York: Norton.

- ERIKSON, ERIK H. 1959 Identity and the Life Cycle: Selected Papers. *Psychological Issues* 1, no. 1. → See especially the "Introduction" written by Rapaport.
- GILL, MERTON M. 1961 David Rapaport: 1911-1960. *Journal of the American Psychoanalytic Association* 9:755-759.
- GILL, MERTON M.; and KLEIN, GEORGE S. 1964 The Structuring of Drive and Reality: David Rapaport's Contributions to Psycho-analysis and Psychology. *International Journal of Psycho-analysis* 45:483-498.
- HARTMANN, HEINZ (1939) 1958 *Ego Psychology and the Problem of Adaptation*. Translated by David Rapaport. New York: International Universities Press. → First published as *Ich-Psychologie und Anpassungsproblem*.
- KNIGHT, ROBERT P. 1961 David Rapaport: 1911-1960. *Psychoanalytic Quarterly* 30:262-264.
- PAUL, IRVING H. 1959 Studies in Remembering: The Reproduction of Connected and Extended Verbal Material. *Psychological Issues* 1, no. 2.
- SCHWARTZ, FRED; and ROUSE, RICHARD O. 1961 The Activation and Recovery of Associations. *Psychological Issues* 3, no. 1.
- WOLFF, PETER H. 1960 The Developmental Psychologies of Jean Piaget and Psychoanalysis. *Psychological Issues* 2, no. 1.

#### RATIONING

See PRICES, article on PRICE CONTROL AND RATIONING.

#### RATZEL, FRIEDRICH

Friedrich Ratzel (1844-1904), German geographer, started his professional life as an apprentice pharmacist but at the age of 21 began seriously to study the natural sciences, especially zoology. After periods of training at several German universities, he received his PH.D. from the University of Heidelberg in 1868 for a dissertation on the oligochaetes, a group of annelids. A year later he published his first book, a popular treatise on the organic world (1869). In 1868-1869 Ratzel traveled in southern France and began a series called "Zoological Letters from the Mediterranean Region," which was published in the *Kölnische Zeitung*. The success of this venture led to his employment as a traveling correspondent. In this capacity he visited eastern and southern Europe before interrupting his journalistic work to serve as a volunteer in the infantry during the Franco-Prussian War. After the war he studied briefly at Munich with the naturalist-ethnographer Moritz Wagner, who exerted a strong influence on his subsequent career. In 1872 Ratzel resumed his work as a traveling correspondent for the *Kölnische Zeitung*, visiting southern Europe, the United States, Mexico, and Cuba. As a result of essays based on these travels and a monograph on Chinese emigration (1876), he was appointed lec-



turer, in 1875, and then professor, in 1876, at the Technische Hochschule in Munich. Ratzel was now well on his way as an academic geographer; in 1886 he was called to the University of Leipzig, where he remained as professor of geography until his death.

Ratzel's creative life can be divided into two periods, that prior to 1876, when he worked as a journalist and perfected his knowledge of biology and geography, and that after 1876, when his work took on a more scholarly and theoretical character. His total output of published work is enormous, consisting of 25 books and 518 shorter works. His first substantial geographic study, a two-volume treatise on the United States (1878-1880), is not markedly different in scope or organization from twentieth-century regional texts. But Ratzel's reputation is not derived primarily from this work or indeed from any of his regional studies. Rather it is based on three systematic treatises, each devoted to a major aspect of human geography. Two of these works, his two-volume *Anthropogeographie* (1882-1891) and his *Politische Geographie* (1897), made him the center of prolonged controversy. The third work, his three-volume *History of Mankind* (1885-1888), outdated as it is, still merits respect as a milestone in the history of ethnography.

The controversial aspect of Ratzel's *Anthropogeographie* results from his tendency, mainly in the first volume of the work, to assert the existence of a direct causal relationship between features of the natural environment and works of man. However, Ratzel's environmentalism was tempered with a historical perspective, and he asserted the man-land relationship with reservations; thus he does not entirely deserve to be labeled an "environmentalist." Indeed, in the English-speaking world this label more properly belongs to Ellen Semple, who studied with Ratzel in the 1890s and published a substantial book based on the first volume of his *Anthropogeographie* (Semple 1911).

The second volume of the *Anthropogeographie* is concerned with the distribution and movement of peoples and cultures, an interpretation of the changing character of the habitable world. The content of this volume, especially its stress on diffusion and migration as major cultural processes, is compatible with more recent conceptions of the essential task of human geography (Steinmetzler 1956).

The controversy over Ratzel's *Politische Geographie* centers on his use of biological analogy in interpreting the expansion or contraction of states. Ratzel emphasized *Raum* (area) and *Lage* (location) as basic elements in the constitution of states

and suggested that such entities might be subject to natural processes of growth and decay. Since states, in this view, function as organisms, they cannot be contained within rigid bounds, and the survival of a nation or culture is linked to its capacity for expansion.

During the Third Reich, Karl Haushofer and his school of Geopolitik carried these suggestions into political ideology, and the doctrine of *lebensraum* (living space) became the pseudoscientific justification for national expansion. However, it is important to realize that Ratzel was not concerned primarily with national policy. A geographer initially trained as a biologist, he simply tried to apply ecological and evolutionary concepts in his studies of political and economic development (Overbeck 1957): for Ratzel, human geography was but one part of a general biogeography, and in trying to formulate principles for this more comprehensive science, he used biological terminology in descriptions of cultural phenomena. Indeed, the concept of *lebensraum*, as applied by Haushofer and his school, was not taken directly from Ratzel's writings, but rather from the derivative work of the Swedish political scientist Kjellén (1916).

The least controversial of Ratzel's major works, *The History of Mankind*, must also be ranked as a major contribution to modern social science. The work is a richly illustrated survey of primitive peoples (*Naturvölker*) on a world-wide scale. Detailed information is offered on the natural environment, racial characteristics, social organization, religion, and material culture of each group described. The primitive condition of the *Naturvölker* is explained primarily by social, technological, and locational factors, and only incidentally by circumstances of the natural environment or by race. Errors of fact and interpretation abound, as one might expect in a work of this scope based on nineteenth-century data. Nevertheless, Ratzel's general ideas remain valuable—"the conception of humanity as a unit, the tempering of environmentalism with a historical perspective, the demand for a conversion of space into time relations, the depreciation of spectacular migrations in favor of slow, continuous infiltration, the postulation of marginal peoples" (Lowie 1937, p. 127). Ratzel's most impressive contribution may well have been this pioneer effort to integrate the methods and concepts of biology, ethnography, and geography.

MARVIN W. MIKESELL

[Other relevant material may be found in ENVIRONMENTALISM; ETHNOGRAPHY; GEOGRAPHY, article on POLITICAL GEOGRAPHY; and in the biography of KJELLÉN.]

## WORKS BY RATZEL

- 1869 *Sein und Werden der organischen Welt: Eine populäre Schöpfungsgeschichte*. Leipzig: Gebhardt & Reisland.
- 1876 *Die chinesische Auswanderung: Ein Beitrag zur Cultur- und Handelsgeographie*. Breslau (Germany): Kern.
- 1879-1880 *Die Vereinigten Staaten von Nord-Amerika*. 2 vols. Munich: Oldenbourg. → Volume 1: *Physikalische Geographie und Naturcharakter der Vereinigten Staaten von Nord-Amerika*. Volume 2: *Kultur-geographie der Vereinigten Staaten von Nord-Amerika*.
- (1882-1891) 1921-1922 *Anthropogeographie*. 2 vols. Stuttgart (Germany): Engelhorn. → Volume 1: *Grundzüge der Anwendung der Erdkunde auf die Geschichte*, 4th ed. Volume 2: *Die geographische Verbreitung des Menschen*, 3d ed.
- (1885-1888) 1896-1898 *The History of Mankind*. 3 vols. London and New York: Macmillan → First published as *Völkerkunde*.
- (1897) 1923 *Politische Geographie*. 3d ed. Edited by Eugen Oberhummer. Munich and Berlin: Oldenbourg.
- 1906 *Kleine Schriften*. 2 vols. Edited by Hans Helmolt, with a bibliography by Viktor Hantzsch. Munich and Berlin: Oldenbourg.

## SUPPLEMENTARY BIBLIOGRAPHY

- HEINE-GELDERN, ROBERT 1964 *One Hundred Years of Ethnological Theory in the German-speaking Countries: Some Milestones*. *Current Anthropology* 5:406-418.
- KJELLÉN, RUDOLF (1916) 1924 *Der Staat als Lebensform*. 4th ed. Berlin: Vowinkel. → First published in Swedish as *Staten som livsform*.
- LOWIE, ROBERT H. 1937 *The History of Ethnological Theory*. New York: Farrar & Rinehart.
- OVERBECK, HERMANN 1957 *Das politischgeographische Lehrgebäude von Friedrich Ratzel in der Sicht unserer Zeit. Die Erde: Zeitschrift der Gesellschaft für Erdkunde zu Berlin* 88:169-192.
- SEMPLE, ELLEN C. 1911 *Influences of Geographic Environment, on the Basis of Ratzel's System of Anthropogeography*. New York: Holt.
- STEINMETZLER, JOHANNES 1956 *Die Anthropogeographie Friedrich Ratzels und ihre ideengeschichtlichen Wurzeln. Bonner geographische Abhandlungen* 19:1-151.
- WANKLYN, HARRIET G. 1961 *Friedrich Ratzel: A Biographical Memoir and Bibliography*. Cambridge Univ. Press.

## RATZENHOFFER, GUSTAV

Toward the end of a successful career in the Austrian army and after retirement, Gustav Ratzenhofer (1842-1904) wrote six books on political science, sociology, and philosophy. He was essentially self-taught in these fields, his formal education having ended after a brief period in secondary school. He had some experience in the family business of clockmaking, but friction with his stepfather led him to leave the business and, in 1859, to become a cadet in the army.

His service as president of the high military court

in Vienna may have stimulated his interest in social science: in a biographical preface to his posthumously published *Soziologie* (1907), his son shows how Ratzenhofer's ideas progressed logically toward sociology and philosophy. An early article focused on some tactical features of a particular campaign; next he wrote on military law; and from there he moved to the consideration of the political context of the law and, finally, to theoretical sociology and ontological and ethical problems. Indeed, sociology was for Ratzenhofer a part of a comprehensive philosophy, which he called positive monism.

Ratzenhofer's political and sociological works are characterized by a preoccupation with the evolution of types of human associations. He reduced social phenomena to physical, chemical, and biological ones and located man's fundamental drives in his biological nature. These drives are self-preservation (rivalry for food, *Brotneid*) and sex (hence the blood bond, *Blutliebe*). Every human being tends to act in terms of these and other basic interests, regardless of opposition, and therefore there exists a state of "absolute hostility" in the interaction of men. Absolute hostility was restrained within the primitive horde by the blood bond and by the felt advantages of cooperation in war and work. Among primitive hordes, the normal relationship was conflict—resulting from increasing contacts between hordes as populations grew. Such conflict often resulted in the conquest of one horde by another and the subsequent formation of a state, in which the exploitation of the defeated constituted the basis of economic and other social activity. Eventually, cultural sanctions developed for what had at first been forced labor by captives (1893, vol. 1, p. 13). Cultural contacts between states, facilitated by commerce, tended to limit the absolute control of a particular state over its subjects. In the long run these processes of limitation resulted in "civilization," the transformation of primitive conquest and exploitation into "the equitable sharing of each individual . . . in the conditions of life" (1893, vol. 3, sec. 61).

Although Ratzenhofer did not acknowledge his debts specifically, his work owes much to Ludwig Gumplowicz, to Herbert Spencer, to Adolf Bastian, and perhaps something to Rudolf von Jhering. Jhering's theory that "interests" are a fundamental factor in law may have been the source of Ratzenhofer's emphasis on interests as the motive force in all human behavior. Ratzenhofer, in turn, had considerable influence on Albion W. Small, whose *General Sociology* (1905) devotes 16 chapters (over two hundred pages) to an interpretation of Ratzenhofer. There seems to be



no specific heritage from Ratzenhofer in the social sciences of today; nevertheless, he was a pioneer in the use of factual realism in political science.

FLOYD N. HOUSE

[For the historical context of Ratzenhofer's work, see the biographies of BASTIAN; GUMPCLOWICZ; SPENCER; for discussion of the subsequent development of Ratzenhofer's ideas, see SOCIAL DARWINISM and the biography of SMALL.]

#### WORKS BY RATZENHOFER

- 1893 *Wesen und Zweck der Politik*. 3 vols. Leipzig: Brockhaus. → Volume 1: *Die sociologische Grundlage*. Volume 2: *Die Staatspolitik nach Aussen*. Volume 3: *Der Zweck der Politik im Allgemeinen*.  
 1898 *Die sociologische Erkenntnis: Positive Philosophie des socialen Lebens*. Leipzig: Brockhaus.  
 1899 *Der positive Monismus und das einheitliche Princip aller Erscheinungen*. Leipzig: Brockhaus.  
 1901 *Positive Ethik: Die Verwirklichung des Sittlich-Seinsollenden*. Leipzig: Brockhaus.  
 1902 *Die Kritik des Intellekts: Positive Erkenntnistheorie*. Leipzig: Brockhaus.  
 (1904) 1906 *The Problems of Sociology*. Volume 5, pages 815-824 in *International Congress of Arts and Science*, St. Louis, 1904, *Congress of Arts and Science*. Edited by Howard J. Rogers. Boston: Houghton Mifflin. → First published in the *American Journal of Sociology*, Volume 10, pages 177-188.  
 1907 *Soziologie: Positive Lehre von den menschlichen Wechselbeziehungen*. Leipzig: Brockhaus.

#### SUPPLEMENTARY BIBLIOGRAPHY

- BENTLEY, ARTHUR F. 1926 Simmel, Durkheim, and Ratzenhofer. *American Journal of Sociology* 32:250-256.  
 GRAMZOW, OTTO 1904 *Gustav Ratzenhofer und seine Philosophie*. Berlin: Schildberger.  
 SMALL, ALBION W. 1905 *General Sociology: An Exposition of the Main Development in Sociological Theory from Spencer to Ratzenhofer*. Univ. of Chicago Press. → See Parts 4 and 5, pages 183-394, for an interpretation of Ratzenhofer's theories.  
 SMALL, ALBION W. 1908 Ratzenhofer's Sociology. *American Journal of Sociology* 13:433-438.

## REACTION TIME

When a human subject follows instructions to make a specific response as soon as he can after the presentation of a specific signal, the latency of the response is called reaction time (RT). Average values of between 150 and 250 milliseconds (msec.) are typically found, for example, where the subject must press a telegraph key when a light is flashed; however, under some conditions, RTs even shorter than one hundred msec. and even longer than one second may occur. The term "RT," is also applied to the continuous lag of responses to the stream of ongoing events.

Beginnings of the RT experiment are found in

two sources. One is Helmholtz' study of the speed of neural transmission. In an unsuccessful attempt, reported in 1850, to adapt the nerve muscle preparation technique to the intact human organism, he measured the latency of voluntary hand responses to light electric shocks applied to different areas of the skin. It may be supposed that Helmholtz did not appreciate what could be learned through this approach as he did not proceed with it. However, it was not long before many others were actively experimenting and theorizing on the problem.

The other source is the astronomer Friedrich Bessel's analysis in 1822 of differences among observers in their estimates of the instant, within a series of clock ticks, in which a star passed a cross wire, the "eye and ear" method. With the interest thus aroused in the variability of human behavior, and with the concurrent development of accurate timing apparatus designed to eliminate this factor from astronomical observations, there eventually resulted the measurement of true RTs by two astronomers. One was O. M. Mitchel, an American, who used light and sound signals in 1858; the other was A. Hirsch, a Swiss, who also used electric shock in 1861-1864.

Three early programs of research on RT are of special importance. The first is that of the Dutch physiologist Frans Donders, who in 1868 introduced the experiment in which response is made dependent upon a choice between signals, i.e., the disjunctive RT. By subtracting the RT to a simpler situation from that to a more complex situation, he hoped to find the time required for the additional mental act. The second is that of Siegmund Exner, who first used the term "reaction time" in 1873. His emphasis was on the importance of preparation and on the subject's feeling of involuntariness of the response. The third is that of Wundt, started in 1880, which was notable for the attempt to measure the RT for the identification of a stimulus. At least the germ of much of the later experimentation on RT and of the accompanying conceptualizations is found in the work of these pioneers. (See Boring [1929] 1950, chapter 8; and Woodworth 1938, chapter 14 for a more complete historical coverage.)

Two characteristics of RT have especially attracted the attention of investigators. First, RT is shortened by "strength" conditions, but not below an irreducible minimum. Second, RT is increased by "complexity" conditions. Much of the research in this field has aimed at developing functional relations between RT and variables concerned with strength or complexity.

**Strength variables.** Considered here are variations which, in some respect, might be expected differentially to activate, facilitate, or energize some part of the chain of processes between signal and response.

**Intensity of signal.** It regularly has been observed that more intense signals have a shorter RT. Moreover, two signals of different frequency but of the same subjective intensity will produce extremely similar distributions of RT. If the signal is the cessation of a stimulus, a shorter RT is again found for a strong stimulus. RT to a sudden change of stimulus intensity is shorter as the amount of change is greater.

**Discriminability.** Discriminability may also be shown to affect RT, and with a wider variety of examples, by the disjunctive, or choice, procedure. The more different the signals, the shorter the RT. This has been demonstrated for frequency of tones, length of lines, gradation of colors, number of objects, and distance between points of tactile stimulation.

**Motivation.** Motivation may, in several ways, be shown to have an effect. For example, informing the subject of his RT after each trial decreases RT. A more marked reduction is found by applying an electric shock for high RTs.

**Preparation.** In most studies of RT a warning signal is given, perhaps a second or two before the signal for response. If the foreperiod (or warning interval) is increased to seven or eight seconds, RT will be lengthened. Too short a warning interval will increase RT, but a greater increase will result from highly variable or unexpected foreperiods (Klemmer 1956). Degree of muscular tension was found to vary inversely with RT under such manipulations.

**Practice.** Practice has been found effective in reducing RT, especially for complex conditions, where considerable reduction may take place over many daily sessions. Warm-up within a session has also been noted, lowest values not ordinarily being attained before the tenth or fifteenth trial.

**Minimum RT and comparison of senses.** A typical subject has a mean auditory RT of about 140 msec. under usual testing conditions and mean visual RT of about 180 msec. The difference is similar to that required to set the respective sensory nerves into action by appropriate stimuli to the sense organs, and the argument has been made that the difference in RT for the two modalities lies in this differential for receptor stimulation. The argument would be more convincing if their irreducible minimums of RT were composed of known processes. However, in the case of auditory

stimulation, only about 55 msec. is accounted for by measurable "peripheral" events and in visual stimulation about 85 msec. (Woodworth [1938] 1960, p. 19). It is risky to assume that, except for the matter of stimulation, the same mechanisms are involved in auditory and visual RT. As a matter of fact, in the study where the lowest RT values were found for the two modalities, the two mean RTs were both close to 110 msec. (Hovland & Bradshaw 1935). In this study the view was presented that the typical difference might be explained by the poorer figure-ground segregation under usual visual testing conditions (or, in other terms, a lower signal-noise ratio), a factor that was eliminated by making the visual signal a bright light in completely dark surroundings. In view of the difference in time for the stimulus process apparently demonstrated for the two modalities, this then would imply, if the subtractive is accepted, that the unknown processes require less time for visual than for auditory signals as the irreducible minimums are approached.

RT for touch is sometimes as short as that for sound, sometimes perhaps 25 msec. longer. There are special difficulties in finding RT for other types of stimulation: taste, smell, pain, etc. One problem is the isolation of specific sense qualities: a sharp point will be felt as pressure as well as pain. Another is to ascertain the instant when the stimulus acts on the receptor, as in the case of olfaction. The RTs reported for these modalities generally range between 300 msec. and one second.

**Complexity conditions.** Complexity is introduced by any circumstance that changes the situation from that of the simple RT, where the same signal appears on each trial, requiring the same response, and in which there is no problem of the relation between signal and response.

**Multiple options.** The RT where the choice of reactions depends upon which of a set of signals appears has been approached from the point of view of information theory. The same increase in RT has been found between two and four options—one and two bits—as between four and eight options—two and three bits (Hyman 1953). However, the effect of number of options is much reduced by extensive practice and by elimination of coding requirements, for example, using a tactual signal to the finger. Also, if options permitting simultaneous response of several fingers are allowed, RT does not increase greatly with an enormous increase in the number of possible patterns.

**Indirectness.** Any introduction of indirectness or incompatibility between signal and response will increase RT. If the lights on a display no



longer directly indicate the required response, *RT* will rise (Garvey & Knowles 1954).

**Prior involvement.** If there is an unpredictable succession of two signals, each requiring a response, *RT* to the second signal is elevated over normal. The shorter the interval between signals, the greater will be the elevation. The explanation of a "psychological refractory period" holds that the processing of a new response cannot start until the end of the processing of the first response (Welford 1959). Alternatively, it has been suggested that the preparation for the second response is interfered with by the first event (Poulton 1950).

**Explanations of *RT*.** It may be well to start with the major problem raised by the relation between intensity of signal and *RT*. Since speed of neural transmission is independent of intensity of stimulus, the relation between intensity and *RT* cannot be explained in terms of delays in transmission along neurons. Four modes of explanation have been proposed (Chocholle 1963).

Two of the modes specify kinds of neurological mechanisms; the other two imply the properties they must have. The first explanation is that there is actually a faster *over-all* neural transmission for stronger signals despite the constant transmission speed along single neurons. One possibility cited is that of more rapid traversal of synapses. This appears inadequate to account for the extent of differences found. Another possibility, that a stronger stimulus obtains access to more direct pathways in the central nervous system (CNS), has not yet been directly tested. The second mode of explanation is that there are sensing mechanisms in the CNS which are activated more quickly by a stronger stimulus because of a higher rate of impulses along ascending neurons or because of the involvement of more fibers. This, too, has yet to be verified.

The third mode of explanation is that *RT* is short to the extent to which part processes between signal and response can be prepared (Poulton 1950). Perhaps full preparation of the effectors (motor set) cannot be attained or maintained if involvement is required of the mechanisms of detection (sensory set).

The final mode of explanation is in terms of a stochastic decision process which distinguishes between a noisy signal and noise alone (McGill 1963). A strong signal might reach a "criterion count" sooner than a weaker one.

The last two modes of explanation lend themselves to aspects of *RT* other than its relation to intensity. With fewer alternatives, there can be greater preparation for those that do occur—especially for those most likely—as well as a "clearing

of the ways" for the response. Also relevant for this expectancy formulation is the relation between *RT* and variability of the length of the foreperiod.

The kind of decision model described explains the relation between *RT* and number of alternatives by the time necessary to distinguish among the "noisy" alternatives. The information-theory approach appears to give a fixed amount of time to each successive binary decision, although the means by which this is accomplished have not been stipulated [see INFORMATION THEORY; see also Bricker 1955].

**Relation to other aspects of behavior.** Reaction time is related to several other areas of investigation.

**Association time.** In the association experiment, the subject makes a verbal response to a stimulus. Unlike in the case of *RT*, he employs some already available association: he reads a word, names an object, gives the opposite of a word, or simply replies to one word with another (Woodworth [1938] 1960, chapter 3).

**Operant conditioning.** In operant conditioning, reinforcement of an operant is first made contingent upon the occurrence of a signal. Next, reinforcement is made to depend upon the rapid appearance of the operant after the signal. Latencies may become very short, in the range of values found for *RT* (Skinner 1946). Thus, experiments with animal subjects can be made to parallel those with human subjects. The relation of latency to signal intensity has already been studied. [See LEARNING, articles on INSTRUMENTAL LEARNING and REINFORCEMENT.]

**Tracking.** In tracking procedures the subject keeps a pointer on a moving target or keeps a target from being moved off center. If the position of the target varies continuously it is difficult—although possible in theory—to determine the *RT* to the appearance of an error. However, the step-tracking task amounts to no more than a variant of the *RT* procedure.

**Instrumentation.** Several procedural aspects are relevant to a discussion of *RT*.

**Signals.** The greatest problem in instrumentation is to make the signal appear substantially instantaneously. Neon lamp and tachistoscopic techniques are superior in this respect to incandescent lamps. Although tones may be introduced instantaneously, a click will be heard at their onset, often contrary to the aims of the experiment.

**Responses.** Manipulation of the telegraph key and the light-pressure snap switch are both satisfactory responses. The response may be either a press or a release. Finger responses are usual, but any movement of the hand, arm, leg, etc. may

be employed. A microphonic switch is the preferred method of recording a vocal response. In the step-tracking procedure, a pointer or lever is generally moved laterally.

**Recording.** An immediately accessible RT after each trial may be obtained by means of a continuously running synchronous clock. The direct-current clutch is engaged with the signal and disengaged with the response. Although such clocks are sufficiently accurate for most purposes and are widely employed, electronic counters are superior in avoiding mechanical problems and in not producing clicks. Moreover, permanent registration may be made by means of print-out devices. In the step-tracking procedure, graphic techniques, often utilizing signals produced by transducers, are used to find an index of the start of the response. It is also possible to use such instruments in conjunction with key pressing and other ungraded responses that provide an early indication of the onset of the response. Muscle action potentials provide an even earlier indication (Bartlett 1963). The delay arising from the mass of the recording element may be avoided by use of a camera technique in conjunction with a cathode-ray tube or by use of magnetic tape recording.

**Special conditions.** For a given condition, where the RT for a three-year-old is 500 msec., an exponential decrease to 150 msec. might be expected to about age thirty, with a shallow increase starting thereafter. For as yet unexplained reasons, slightly lower RTs are found for males than for females.

Predictive use has been made in a number of ways. In testing or training automobile drivers a measure is made of time taken to apply the brake on the appearance of a danger signal. Some negative correlations have been revealed between RT and measures of athletic performance.

A variety of special conditions have been tested; following are a few examples. Prolonged vigilance has been found to increase RT, but no consistent effect has been found to result from simple loss of sleep. Benzedrine can overcome the effect of prolonged vigilance but otherwise has no discernible effect. Alcohol has been found to increase both the mean and the variability of RT.

ROBERT M. GOTTSANKER

[Other relevant material may be found in ATTENTION; HEARING; NERVOUS SYSTEM; SENSES; VISION.]

#### BIBLIOGRAPHY

BARTLETT, NEIL R. 1963 A Comparison of Manual Reaction Times as Measured by Three Sensitive Indices. *Psychological Record* 13:51-56.

- BORING, EDWIN G. (1929) 1950 *A History of Experimental Psychology*. 2d ed. New York: Appleton.
- BRICKER, PETER D. 1955 Information Measurement and Reaction Time: A Review. Pages 350-359 in Henry Quastler (editor), *Information Theory in Psychology*. Glencoe, Ill.: Free Press.
- CHOCOLLE, RENÉ 1945 Variations des temps de réaction auditifs en fonction de l'intensité à diverses fréquences. *Année psychologique* 41/42:65-124.
- CHOCOLLE, RENÉ 1963 Les temps de réaction. Volume 2, pages 63-112 in Paul Fraisse and Jean Piaget (editors), *Traité de psychologie expérimentale*. Paris: Presses Universitaires de France.
- GARVEY, W. D.; and KNOWLES, W. B. 1954 Response Time Patterns Associated With Various Display-Control Relationships. *Journal of Experimental Psychology* 47:315-322.
- HOVLAND, CARL I.; and BRADSHAW, DOROTHY A. 1935 Visual Reaction Time as a Function of Stimulus Background Contrast. *Psychologische Forschung* 21: 50-55.
- HYMAN, RAY 1953 Stimulus Information as a Determinant of Reaction Time. *Journal of Experimental Psychology* 45:188-196.
- KLEMMER, EDMUND T. 1956 Time Uncertainty in Simple Reaction Time. *Journal of Experimental Psychology* 51:179-184.
- McGILL, WILLIAM J. 1963 Stochastic Latency Mechanisms. Volume 1, pages 309-360 in R. Duncan Luce, Robert R. Bush, and Eugene Galanter (editors), *Handbook of Mathematical Psychology*. New York: Wiley.
- POULTON, E. C. 1950 Perceptual Anticipation and Reaction Time. *Quarterly Journal of Experimental Psychology* 2:99-118.
- SKINNER, B. F. 1946 Differential Reinforcement With Respect to Time. *American Psychologist* 1:274-275.
- TEICHER, WARREN H. 1954 Recent Studies of Simple Reaction Time. *Psychological Bulletin* 51:128-149.
- WELFORD, A. T. 1959 Evidence of a Single-channel Decision Mechanism Limiting Performance in a Serial Reaction Task. *Quarterly Journal of Experimental Psychology* 11:193-210.
- WOODWORTH, ROBERT S. (1938) 1960 *Experimental Psychology*. Rev. ed. by Robert S. Woodworth and Harold Schlosberg. New York: Holt.

#### READING DISABILITIES

The failure of some children to learn to read as easily or in the same manner as their peers has been a persistent educational problem that has caused much heartache to the children themselves, attracted much attention from research workers, and provoked much controversy in educational circles.

Estimates of the number of children who have severe difficulty in learning to read vary from country to country and from investigator to investigator, depending largely upon the way that disability is defined. In general, such estimates run from 5 to 20 per cent. Boys outnumber girls three to one. Although fewer difficulties are reported in countries where the language is phonetic than in countries where the language is less regular,



even in such countries there persists a small hard core of children with difficulties. Accurate estimates of incidence are often clouded by inadequate diagnosis; children who have reading difficulty may be considered merely unintelligent or careless workers.

It is customary in a discussion of reading difficulty to exclude those children who are genuinely mentally retarded, who are partially sighted, hard of hearing, have gross neurological impairment, or are psychotic. Obviously, differences in these areas are of degree, and distinctions are difficult to make. Valuable insights for the teaching of less-handicapped children have come from work with the severely handicapped.

### Types of reading difficulty

It is possible to identify a number of different types of reading difficulty, although there is much overlap among the types.

**"Simple" reading difficulty.** First, there is what may be called the simple type. This appears in a child who is of normal intelligence, who has no serious physical or emotional problems, whose home background is reasonably stable, and whose background experiences enable him to comprehend the ideas presented in school but who may have slight difficulty in one or more of these areas and who may have received instruction that was inadequate or interrupted by illness or by changes of schools, teachers, or methods of teaching. If, for example, a child has only low average ability, or has a minor visual problem, or although not seriously disturbed is timid and feels incompetent, or comes from a home that, although stable, does not value education, or has had only minimal experiences that help him understand his reading, or has any other slight but significant handicap, he may encounter trouble unless he has had superior instruction in reading from the beginning. It is almost as if he can carry a reasonable burden of stress until there comes "the straw that breaks the camel's back." To call this type of difficulty simple is perhaps misleading; with its intricately interwoven threads it is highly complex. But it is possible to unravel the threads with relative ease.

**Experiential inadequacies.** A second type of difficulty, which may be called experiential, has its roots in the differences between school and home. When a child comes from a home that in its values or its patterns of language and behavior is markedly different from the school, he is likely to have difficulty in school in general and in reading in particular.

**Language differences.** Many children do not speak the language spoken at the school they enter.

They may speak a truly different language, a foreign or a tribal language, or they may speak a vernacular or dialectal form of the school language. Their ears may not be tuned to the speech patterns of the teacher; they may not be able to identify the sounds used for the letters. They may not, for example, be able to discriminate between the English sounds "th" and "d." If many of the children face the same problem and if the teacher takes time to lay a foundation in the spoken language before moving to a formal reading program all may be well. But if the child, who does not even realize his problem, is left to flounder he may fall farther and farther behind. Slow reading even at the college level may spring from a persistent unfamiliarity with the language patterns [see LANGUAGE, article on SPEECH PATHOLOGY; PERCEPTION, article on SPEECH PERCEPTION].

**Lack of preparation for reading tasks.** In addition to his problem with the language of the school the child may face tasks and ideas for which his experience has not prepared him. He may have had few stories told or read to him and may not have skill in following a plot. He may have lived a very restricted and isolated life and not know through experience what, for example, a city, or lake, a mountain, or a clown is like. True, he will ultimately use books to expand his horizons, but in the beginning stages of reading he needs to know concretely the referents for which words stand. This lack of a foundation of concrete experiences on which to build the ideas met in reading handicaps many learners. This handicap, too, may persist for many years and contribute to the poor comprehension of many high school students.

**Lack of familiarity with school values.** Unfamiliarity with the values of the school may contribute to reading difficulty. Sitting still, carrying tasks through to completion, talking about events and one's reactions, listening to others—all are skills the child from certain kinds of homes may not have. Certainly he will be handicapped.

**Emotional pressures.** A third type of difficulty springs from the emotional pressures under which the child may be living. If he has fears that he will fail, that he is no good, that he will lose his parents, that his siblings are more competent, or that fate is unjust, his energies will be unavailable for learning. Sometimes reading may represent one area in which a youngster may be master; no one can make him be a successful reader. These and other disturbing feelings may be the major contributing factor to the difficulties of certain children.

**Neurological dysfunction.** Still another type of reading difficulty can be identified—one that has

as its source neurological dysfunction. This type of difficulty has been called dyslexia, word-blindness, and primary reading difficulty. Often children suffering from this type of difficulty are divided into two categories: those whose history and neurological examination show brain injury, however minimal the impairment; and those whose history and examination are essentially negative but whose generalized language problems presuppose some kind of neurological dysfunction. In general, children in both categories may test in the normal or even superior range on an intelligence test but may show marked discrepancies between subtest scores. They have particular difficulty with perceptual-motor tasks, such as copying geometric figures, and with concepts of time, space, and distance.

The academic performance of these youngsters is often highly erratic; their work may be excellent in mathematics and poor in spelling and reading. They may have been late talkers or have some history of speech difficulty. They may have poor auditory discrimination. Their handwriting is often slow, labored, and inferior. They may be awkward, poorly coordinated, and lack manual dexterity. Sometimes, although not always, they have trouble with direction, confusing right and left. They are frequently hyperactive, impulsive, and distractible. Perseveration may mark their performance. The children in the second category show fewer of the above difficulties than do those where the damage is identifiable. Indeed, their difficulties may be restricted to the language areas with, possibly, some directional difficulty. All children do not show all the above characteristics.

**Effects of types of difficulties.** The reading performance of the child reflects the type and source of the difficulty, whether poor instruction, experiential inadequacy for reading, emotional tension, or neurological dysfunction. The reading of the child with a "simple" reading difficulty may reveal the gaps in his instruction, such as having missed many days of school in the first grade. He may, for example, know a good deal about syllabication but not know the sounds of the initial consonants; or his reading errors may show no specific pattern but his general performance be like that of a younger child.

The reader whose major difficulty comes from lack of experience or from a discrepancy between the values of home and school is likely to have particular difficulty with comprehension, although he will probably have difficulty with new words as well and may appear uninterested in books and stories. His general academic achievement in other subjects, too, is likely to be poor.

Emotional tensions reveal themselves in many ways: in inability to concentrate; erratic test performance; discrepancy between test performance and school achievement; slow, inflexible, overly precise approach to all reading tasks; blocking at certain words; and in an apathetic, bored tone or an anxious, tense one. Any of these difficulties may, of course, have causes other than emotional tension.

The child whose difficulty is of neurological origin may show a variety of reading characteristics. He may reverse words or letters in reading and spelling, writing *lithg* for *light* or reading *on* for *no*. He often guesses at words on the basis of first letter. He can discriminate between two similar words if both are in front of him, but he cannot retain a visual image of a word. Vowels and the blending of letter sounds often give him difficulty. Most frustrating to his teacher is his tendency to know a word on one line and miss it on the next. Surely he is just careless! The sloppy nature of his written work adds to this picture. "He could do it if he tried," claims the parent or teacher who does not understand his problem.

### Diagnosis

It is easier to identify in theory the sources of difficulty and to differentiate between the types of behavior and performance shown than it is to make a differential diagnosis in practice. There is much overlap among types, and the diagnostic tools are crude. An adequate developmental history is of the greatest importance. School reports and records and tests of vision, hearing, and intelligence are useful. Tests to identify perceptual-motor difficulties or memory defects are not readily available or highly refined. Often the clinician must rely on subjective impressions and on informal appraisals. Much improvement in these measuring devices can be expected in the next few years.

Diagnosis of reading difficulty involves a survey of the child's general level of reading skill and of his specific strengths and weaknesses in reading. The diagnostician wants to know how much and what kind of reading he does, whether he knows the phonetic elements of the language, whether he can and does use this knowledge to decipher unknown words. It is important to know, too, his reading speed and his competence in such comprehension skills as skimming, getting the main idea, or picking out the important details. An analysis of the errors he makes often reveals gaps in his knowledge and sheds light on his mode of attack. Among older children, a comparison of performance on timed and untimed tests will reveal the painfully slow but accurate reader and



the one who cannot understand the material no matter how much time he has. Observation and examination of the child's errors are as important as test scores in appraising performance.

### Treatment

When the nature of the child's difficulties and their probable sources have been determined, the method of treatment then needs to be considered. Too frequently some one remedial technique has been advocated almost as a "magic" cure-all. This search for a panacea may be touching in its naiveté, but it has been fruitless and even dangerous. There is no one form of treatment that will help all children with difficulty. Too often children have had years of remedial help with few or no positive results. This is not because of the child's inability to learn; even seriously handicapped children can make slow but steady progress. Too often it is because the techniques and materials have been chosen with little regard for the child's needs as revealed by his performance and behavior.

Since most children with reading difficulty have met much failure and since the acquisition of reading skill is an important developmental task of childhood, one of the most important elements in treatment is success. The task must be broken down into such small steps and the child given so many clues to follow that he can hardly fail to respond correctly. False praise does not help him; real success does. All the teacher knows about human learning and about pedagogy will be of service in a remedial program.

**Specific approaches.** Beyond the application of general techniques of teaching, certain specific approaches seem effective with certain types of difficulty. For the child whose difficulty springs from poor or inadequate instruction, careful pinpointing of his gaps in knowledge, followed by systematic instruction with interesting material, is usually sufficient. For the child whose experiences are limited and whose motivation for learning is insufficient, it is often necessary to provide evidence that reading skill is useful. Helping him to read material that will enable him to do something he wants to do, for example, make a model plane or a batch of cookies, may spark his interest. It is wise to avoid abstract, dull practice material that savors too much of school, in which he has failed and the value of which he often does not see. Allowing him to dictate stories of his own experiences and using these as the basis of instruction has also been found useful. Practice in all forms of language usage—listening and talking as well as reading and writing—is essential.

**Psychotherapy.** For the child whose difficulty springs from emotional tension a variety of techniques have been tried. Some form of psychotherapy, preceding or concurrent with special reading instruction, is often effective. Too often, however, psychotherapists have urged remedial teaching for a child before he is emotionally ready to profit from it. Such unwise haste often adds another failure to the child's already crushing burden. Conversely, some psychotherapists have acted as if the child would be able to read as soon as he felt better about himself and others. Sometimes "magic" does seem to happen, but more often psychotherapy only makes it possible for the child to use his energies in learning—he must still be taught.

**Treatment of the dyslexic.** Of all types of difficulty, those springing from neurological dysfunction require the most carefully planned treatment. For the dyslexic child, especially, remedial treatment that does not work is harmful, leading to confusion and distraction. It is fortunate that many of the methods needed to work with the dyslexic youngster are also useful with the child whose problems stem from emotional pressure. The latter child will need some supplementary enrichment of the remedial program and can better tolerate violations of the general principles of learning.

In treatment of the dyslexic it is essential to keep in mind the nature of his difficulties. He has difficulty with perception; he therefore needs instruction that will focus on the specific perceptual task. He needs to have his attention drawn to the details of the word. He needs much practice in noting the differences between words; for example, in differentiating between *get* and *got*. Since he may have difficulty with auditory discrimination as well as with visual perception it is often helpful to use a multisensory approach, having him write or trace the word as well as say it. Such practice also focuses on details.

The dyslexic tends to repeat his mistakes; it is important for the teacher to set up situations in which he does not make mistakes. This is done, in part, by dividing the task into very small elements and presenting these so gradually that it is difficult to make a mistake. It is also done by providing a word for him as soon as he hesitates. The dyslexic child needs to be able to use his phonic skills rapidly and automatically. Slow, laborious sounding out defeats its purpose.

Since the dyslexic child is easily distracted and confused, materials should be simple, uncluttered, and based on skills he already has. Encouraging

the dyslexic child to jump into materials that are much too difficult for him is disastrous. He does not swim; he sinks. He seems to lose even skills he previously had. In addition to his difficulty in reading, the dyslexic also often has trouble with spelling and handwriting. Too many written exercises based on his reading may be discouraging.

**Treatment programs.** Instruction tailored to the child's type of difficulty and specific needs can be provided under several organizational patterns. One fruitful practice involves the early identification of children who may have trouble and the planning for them of a preventive program. This has been done for the experientially deprived and for the potentially dyslexic. Such a program has the obvious advantage of eliminating the depressing failure so often experienced by these children.

Individual instruction, perhaps over a period of years, is the most effective pattern of help for dyslexics. This can be done within the regular school setting or outside the school in a clinic or by a private practitioner. At the beginning stages these children rarely function well in a group; hence, even in a classroom teaching must be individualized.

The most widespread pattern of treatment for all children with reading difficulty has been the small remedial group, meeting three to four times a week in school but outside the regular classroom. For children with gaps in their past instruction and for those of deprived cultural background, and for some emotionally troubled children, such groups have been effective if the group was small enough, the instruction intensive, coordinated with classroom work, and continued for a long enough period of time, and the teacher sensitive and flexible. When these conditions have not been met, and when dyslexics and seriously disturbed children have been included, the remedial group has been minimally effective. A good program of treatment provides for both individual and group help as each seems appropriate. Besides remedial treatment provided in schools, a number of social agencies, courts, and individuals offer supplementary services.

**Materials used in treatment.** Materials used in the treatment of reading difficulty are many and varied, ranging from teacher-made games and exercises to rather elaborate instruments providing tachistoscopic training. The materials themselves are of less importance than the way they are used. There is no magic in one kind of exercise, one type of vocabulary control, one organization of skills. The secret of successful use of materials lies in selecting those appropriate to a certain

child with a certain type of difficulty. The colorful, delightfully illustrated, amusing book that catches the attention of the culturally deprived ten-year-old may only confuse, distract, and overwhelm the child with slight neurological damage. The phonic exercise book may provide structure for a disorganized child but bore or defeat one who dislikes school. Machines that compel the reader to proceed more rapidly than usual may goad one college student on to more rapid reading and frustrate and annoy another. Appropriateness is the key to successful selection of materials.

A technical society cannot afford to have a group of unemployable, inadequate readers. As the pressure for trained manpower increases, increased effort must be made to provide treatment for all children who have difficulty in learning to read.

ANNE MCKILLOP

[Other relevant material may be found in DEVELOPMENTAL PSYCHOLOGY; INTELLIGENCE AND INTELLIGENCE TESTING; LANGUAGE, article on LANGUAGE DEVELOPMENT; LEARNING, article on NEUROPHYSIOLOGICAL ASPECTS; MENTAL RETARDATION.]

#### BIBLIOGRAPHY

- BOND, GUY L.; and TINKER, MILES A. 1957 *Reading Difficulties: Their Diagnosis and Correction*. New York: Appleton.
- BRYANT, N. DALE 1962 Reading Disability: Part of a Syndrome of Neurological Dysfunctioning. Pages 139-143 in International Reading Association, *Conference Proceedings Volume 7: Challenge and Experiment in Reading*. Edited by J. Allen Figurel. New York: Scholastic Magazine Press.
- DEHIRSCH, KATRINA 1961 Reading and Total Language Disability. Pages 211-214 in International Reading Association, *Conference Proceedings Volume 6: Changing Concepts of Reading Instruction*. Edited by J. Allen Figurel. New York: Scholastic Magazine Press.
- EPHRON, BEULAH 1953 *Emotional Difficulties in Reading*. New York: Julian Press.
- GIBSON, ELEANOR J. 1965 Learning to Read. *Science* 148: 1066-1072.
- HARRIS, ALBERT J. (1940) 1961 *How to Increase Reading Ability*. 4th ed. New York: McKay.
- HARRIS, IRVING D. 1961 *Emotional Blocks to Learning*. New York: Free Press.
- JOHNS HOPKINS CONFERENCE ON RESEARCH NEEDS AND PROSPECTS 1962 *Reading Disability: Progress and Research Needs in Dyslexia*. Baltimore: Johns Hopkins Press.
- KAWI, ALI A.; and PASAMANICK, BENJAMIN 1959 *Prenatal and Perinatal Factors in the Development of Childhood Reading Disorders*. Monographs of the Society for Research in Child Development, Vol. 24. Lafayette, Ind.: Child Development Publications.
- PEARSON, GERALD 1954 *Psychoanalysis and the Education of the Child*. Part 2: Ego Psychology and the Education of the Child. New York: Norton.
- RABINOVITCH, RALPH L. 1959 Reading and Learning Disabilities. Volume 1, pages 857-869 in Silvano



- Arieti (editor), *American Handbook of Psychiatry*. New York: Basic Books.
- RIESSMAN, FRANK 1962 *The Culturally Deprived Child*. New York: Harper.
- ROSWELL, FLORENCE; and NATCHEZ, GLADYS 1964 *Reading Disability: Diagnosis and Treatment*. New York: Basic Books.

## REASON ANALYSIS

Reason analysis is a set of procedures used in survey research to construct causal explanations for the actions, decisions, or intentions of individuals (Lazarsfeld 1935). It involves framing and asking questions, as well as coding and analyzing replies of respondents. Typical actions that have been subjected to reason analysis include buying a product (Kornhauser & Lazarsfeld 1935); voting (Gaudet 1955); choosing an occupation (Lazarsfeld 1931); becoming a juvenile delinquent (Burt 1925); getting married or divorced (Goode 1956); joining a voluntary association (Sills 1957); moving from one home to another (Rossi 1955); and going to a psychoanalyst (Kadushin 1958; 1962; 1966). Occasionally, reasons for *not* acting, such as not using contraception or other family-planning procedures, are also studied (Sills 1961).

**Assessment of cause.** The essential difference between reason analysis and other forms of survey analysis is the method used for assessment of cause. Suppose that in a cross-sectional survey 55 per cent of those who listened to a campaign speech voted for a candidate; but of those who did not listen to him, only 45 per cent voted for him. Thus, the speech might have influenced 10 per cent of the electorate to vote for the candidate. In contrast, a reason analysis is not cross sectional, for it deals either with those who have acted or, alternatively, not acted; or, if both actors and non-actors are present (Berelson et al. 1954; Michigan 1960), it separately analyzes the reasons given by actors and nonactors. In a reason analysis, we might decide through questioning of the voters that 10 per cent voted for the candidate only because they had heard his speech (Lazarsfeld 1942). There are thus 10 per cent who may be said to have voted for the candidate "because" of the speech in both examples, but the meaning is quite different. In a cross-sectional survey the 10 per cent represents the net difference over the entire sample and is dependent on a logic similar to that of a controlled experiment, whereas in a reason analysis the 10 per cent represents the sum of the evaluations of the actions of individual voters.

Can such evaluations be valid? Any causal

variate must precede or logically encompass the dependent variate (the one that it is said to "cause") and must be linked to it through certain mechanisms. Because the respondents in cross-sectional surveys are not randomly assigned to states of the causal variate (Selvin 1957), it is possible, in principle, for the observed relationship to be a spurious one. In the example cited above, we know that persons more favorably disposed to a candidate are somewhat more likely to listen to his speeches to begin with (Berelson et al. 1954), and so the speech undoubtedly did not have a full 10 per cent effect. Theory determines whether such variates are likely to be found, and it also indicates the mechanisms, or intervening variates, that link the causal to the dependent variate. Reason analysis does not have the same statistical criteria of causal relationship that cross-sectional surveys have (Lazarsfeld & Rosenberg 1955, p. 389), and it is therefore even more dependent on theory. To say that an actor acted because of a given factor is to say that other factors were not as important. Reason analyses generally fail if they do not begin with a model of behavior that specifies all the relevant factors that might impel an action or prevent one.

**When to apply reason analysis.** Reason analysis can always be used in studying the subjective factors in any course of individual action. In addition, reason analysis is the method to be preferred if one or more of the following conditions hold true: a process is being studied; the act is extremely frequent or extremely infrequent; only those performing the act can conveniently be located or followed. It is therefore sometimes used for studying nonsubjective acts, such as automobile accidents (Haddon et al. 1964; Suchman 1961).

If one wants to know how an action came to be—what steps were taken and what the key choices were; what the actor thought he was doing and how he felt about it; what influences were present and what triggered the action; and, finally, what outcomes the actor expected—then no technique other than reason analysis can be used. One cannot meaningfully ask persons who did not apply to psychiatrists when they decided their problems were *not* severe enough to merit psychiatric attention; meaningful "why not" questions apply only to situations in which there is a social expectation of action. Moreover, even if actors and nonactors are asked their reasons, the format of reasons is rarely identical.

**Frequency of action.** If acts are either extremely frequent or relatively rare, comparisons with nonactors are expensive and perhaps not even meaningful, since the exceptions may be so differ-

ent from the rule as to make comparison irrelevant. Almost everybody has an occupation; hence occupational choice can profitably be studied through reason analysis (Lazarsfeld 1931). On the other hand, since very few go to psychiatrists, a random sample turns up too few cases to work with. Thus, one survey (Gurin et al. 1960) found that only 42 of 2,450 Americans over 21 and living in private households had been to a psychiatrist. Since the criteria for pseudoexperimental matching (Zetterberg 1954, pp. 145–148) of those who did go to a psychiatrist against those who did not are not clear, reason analysis is the method to be preferred in these and similar circumstances.

**Convenience.** Reason analysis is often performed simply because it is administratively convenient. Those who study the membership of an organization, the readers of a library, or other such populations may wish to find out why such persons joined the organization or came to use the facilities. Rather than embark on a new study of members versus nonmembers, they may use reason analysis and thus infer the reasons for joining.

The same applies to studies of long-term processes. While multiwave panel analysis is useful for such purposes and can actually be combined with reason analysis, the latter, since it is a retrospective method, is more easily suited to the study of long-term decisions.

### Models of action and accounting schemes

A model of action—that is, a list of the basic elements in terms of which human action can be described, together with some notion of how action proceeds—is essential to the development of an “accounting scheme.” Such a scheme contains an organized list of all the *factors* that, for the specific purposes at hand, are said to produce or inhibit an action. An element of action is one in a broad category of factors that propel or repel. For example, a craving for ice cream may be a factor in an accounting scheme for studying ice cream purchases; being a six-year-old boy is an important characteristic of an actor that must be taken into account but is not *in itself* a propelling factor, even though a cross-sectional analysis might show high ice cream consumption among six-year-olds as compared to sixty-year-olds.

All current models of action used in reason analysis emphasize the interplay between the subjective point of view and needs of the actor, on the one hand, and the restraints, requirements, and stimulation of his environment, on the other (Hinkle 1963). In this respect these models are similar to those used by phenomenologists (Tiryakian 1965), although sociologists guided by this

point of view have generally not undertaken surveys. The following are some models of action that have been used to guide the construction of accounting schemes.

Kornhauser and Lazarsfeld (1935), in reviewing early studies of buying, suggested that the elements of these actions could be partitioned into two categories: the *individual*—his motive and other mechanisms, such as the state of his knowledge; and the *situation*—in this case consisting of the product and the way it is sold, as well as various influences upon the individual.

Parsons' theory of action offers a similar set of elements: the actor, his ends, the situation, the means and conditions in a situation, and the norms that govern the selection of means; all the elements in this model “appear from the point of view of the actor whose action is being analyzed” (1937, p. 44). In a later version Parsons and his colleagues (Parsons & Shils 1951) substituted, for ends, means, conditions, and norms, the notion of motivational and valuational orientations. The motivational orientation consists of cognitions (what an actor sees), cathexes (what emotional meaning he invests in objects), and evaluations (what weight he gives to objects). The valuational orientation sets standards for cognitions, cathexes, and evaluations. However, both of Parsons' schemes have been used more frequently for developing typologies of actions and institutional norms than as guides to the analysis of concrete actions.

Bales's interaction categories (1950), developed from studies of small groups, appear more relevant to actions in which other actors, seen face to face, are the only “objects”; this situation is not usually encountered in most areas where reason analysis has been applied. Parsons and Bales saw their schemes as related, however, and together developed another scheme (Parsons et al. 1953) that introduced a time dimension to Parsons' model. Here action moves through four phases: adaptation, goal achievement, integration, and latency [see INTERACTION, *article on INTERACTION PROCESS ANALYSIS*].

Since these models are only general guides to the elements that must be included in an accounting scheme, the reason analysis must specify the precise factors to be studied. For example, a reason analysis of soap buying would include exposure to advertising among its “pushes” in the “situation”; a voting study would, instead, include campaign messages; and a study of choice of psychiatrist would check on referral sources.

The selection of factors depends in part on the type of action to be studied. First, reason analysis is usually concerned with acts that involve some



sort of conscious decision; habitual acts are probably not suited for any of these models. A second distinction is that between "depth" acts (i.e., ones which involve a change in self concept) and acts which are more casual. Depth acts are typically more painful for the actor and involve a longer time span: getting married, choosing an occupation, or going to a psychotherapist are all actions that involve a change in self concept. In contrast, buying groceries and even voting are more casual acts, which do not have such consequences for the individual. Finally, there are instrumental decisions or actions, such as entering business, buying stocks, and the like; these seem to involve more cognitive and more self-consciously rational elements than other types of decisions or actions.

The time perspective of an act determines how many separate decisions, acts, or phases must be studied to understand a given action, as well as the class of factors that enters at each point in the action. Thus, each action or decision is divided into appropriate elements and factors. A depth decision, such as going to a psychiatrist, may involve many steps: realizing one has a problem, talking to friends, reading books, deciding to see a professional and choosing a particular one. Casual acts, such as soap buying, might be divided into only two steps: predisposing and precipitating factors. Further, depth decisions tend to be harder to reverse than casual ones. The key time points to study in irreversible acts are obviously those that tend to limit further alternatives or that serve as prerequisites for the next act. In occupational choice, for example, the decision to attend a technical or an academic high school is a key to the range of occupations that a student may then consider.

### Designing a reason analysis

Reason analysis itself is a complex act and thus has several stages. First, types of action involved in the subject to be studied are distinguished one from another; second, the act is divided into phases or separate acts, if this is necessary; third, an accounting scheme is developed for each act or phase; fourth, the accounting scheme is translated into a data-collection guide, which is typically an interview schedule; fifth, a calculus of factors must be developed so that the relative weight of different factors can be assessed. Finally, the results of this assessment are tabulated for the sample as a whole or for different segments of it.

These last two stages can only arise if the first four have been followed. Unless respondents are queried about each of the relevant factors, the

multiple causation of an act is not apparent. Mixing several factors belonging to different elements and then setting their total equal to 100 per cent is a typical error. Another error is mixing several factors belonging to several elements into a single list, which is then factor analyzed. This method does not give the process of an action. The "dimensions" discovered may represent merely the elements of an action that should have been conceptualized to begin with. These and other similar errors have led responsible researchers to conclude that reasons are inadequate data for scientific inquiry. The fault lies, not in the reasons, but rather in the way they are collected and tabulated.

**Interviews and questionnaires.** Since an accounting scheme is the researcher's own miniature theory of action, rather than a questionnaire or interview schedule, questions about the various elements of an action must be written as a usable "script," fitted to the respondent's frame of reference. Unspecified "why" questions reveal the respondent's frame of reference. Therefore, beginning an interview with a "why" question allows the respondent to indicate what is problematic about a given situation. Afterward the original "why" question is dropped.

Most questions in the reason analysis "script" attempt to elicit answers that specify and verify the relationship between the actor's own experience and the structure of his situation. For example, a respondent says he buys gasoline from the nearest station. Since the interviewer has an actual map of the neighborhood, probing may reveal that an even nearer station is not patronized because it does not sell a national brand—in this case, the accounting scheme succeeds in revealing an implicit norm. After the respondent has had his initial say, it is often useful to follow with a checklist, determined by the accounting scheme, to insure that aspects not immediately relevant or accessible to the respondent will be covered.

Questions on the timing of influences or precipitants help to unravel the structure of an action. These questions also allow the researcher to separate a sequence of acts into smaller-unit acts. As a rule of thumb, only those actions that a respondent can talk about as a whole, without skipping from one locus of events to another, are unit acts.

Most reason analysis interviews skip about a good deal until all dimensions are revealed through probing. Thus, experienced interviewers are usually necessary. On the other hand, if an act has been very carefully partitioned into units, elements, and factors; if pretests reveal a complete range of what

respondents are likely to say; and if the respondents are sufficiently motivated to answer some open-ended questions, then it is possible to obtain from self-administered questionnaires results that are comparable to those obtained from interviews.

### Strategies for assessing cause

The most difficult part of a reason analysis occurs after the data have been collected: assessment of cause and meaning must be performed. Since only actors (or only nonactors) are present in the data, cause cannot be assessed by comparing actors to nonactors and noting the differences between them. Other analytic devices must be used. Three strategies for analyzing factors directly revealed by the actors are getting the actor to do so; using the clinical judgment of the researchers; mechanically reducing a number of factors into a smaller set.

In simple actions, with limited accounting schemes, the actor himself may effectively make the assessment as to which element was most important in his action. This procedure is employed after inquiry has already revealed to both actor and researcher all the elements indicated by the accounting scheme. Empirical tests have shown that this method is surprisingly valid, much as the best way of discovering an actor's interest in an election is not to make a complex index but merely to ask him how interested he is (Katz & Lazarsfeld 1955, p. 171; Rossi 1955, pp. 136-137). Of course, the actor's own opinion of his most important reason for acting is a valuable datum even if it is not objectively correct.

The second technique is much like the first, except that in this case the researcher or a panel of judges performs the same evaluation (Lazarsfeld & Rosenberg 1955, pp. 401-409). In this assessment the researcher performs a task similar to that of historians who construct the "cause" of a historical event. Nagel (1961) suggested that historians implicitly invoke a series of logical strategies, both to assess causes and to rank them in order of importance. Such strategies can also be applied to reason analysis and, for that matter, to any "clinical" judgment of cause.

In the third technique, the "property space" (Barton 1955) formed by the various factors in the action is reduced in some mechanical fashion (Sills 1957; Kadushin 1958). If we knew for each voter (in a grossly oversimplified model of voting) that personal influence, campaign speeches, issues, and party loyalty were or were not factors in his choice, we could form a 16-cell property space (2<sup>4</sup>). Reduction consists of combining the cells

according to some rule. Cells are combined when elements or factors are combined or when, under certain conditions, the cells themselves are grouped. We might decide, in the example just given, that party loyalty would be a factor only if the other factors were not present or, that if personal influence were present, all other factors would be ignored. In contrast to the first two methods, the grounds for the reduction are clear and mechanical—an advantage in large-scale research. On the other hand, no exceptions for the intricacies of an individual case are allowed. Although reduction could, in principle, be performed *before* the collected cases are tabulated, almost all property spaces are reduced *after* tabulation, since the way cases cluster together on the various factors often determines how the cells will be collapsed. The meaning of these first three methods for assessing causes is inherent in the factors themselves. Statistical tabulation gives merely the total result of classifying each case separately.

In the following two methods the meaning of the causes can only be induced *after* tabulation. The fourth method combines statistical analysis with clinical assessment in order to induce the relative "impact" of the elements of an accounting scheme (Lazarsfeld 1942). Those actors who are judged to have acted mainly because of a given factor are compared with all of the actors who claimed to be exposed to the factor. If, *among buyers*, 20 of the 100 persons who saw an advertisement bought the product because of the advertisement, then the "impact" of the ad is 20 per cent. This can be compared to the impact of personal influence or to other factors in an accounting scheme. However, as will be shown, to close the argument completely one also needs to know how many *nonbuyers* saw the ad.

The fifth method is exactly like the usual survey analysis techniques and is called morphological analysis (Lazarsfeld 1959). It consists of using a factor or combination of factors to create *types* of actors or *degrees* of action. Further, in a multiple-stage action, or one with many choice points, not all actors have gone through the same stages; different paths may have led to the same outcome. All these differences can be exploited to produce analysis similar to that which would be undertaken if both actors and nonactors were present, except that the tables now concern various types of actors and their differing reasons for acting. Inferences drawn from these differences may aid the understanding of why they came to perform their actions. Morphological analysis can, of course, be combined with any of the first four methods mentioned



above (Katz & Lazarsfeld 1955, p. 171). For example, reasons for action at one stage may be compared with reasons at another stage (Mills et al. 1950).

**Locating the assessment.** Despite its ability to locate explicit intervening variables and mechanisms in action, reason analysis does have some inherent limitations because it deals only with actors. Careful logic, theory, and substantive knowledge can, however, lead the analyst to powerful conclusions. This logical-empirical network of reasoning can be considerably enhanced by tying it at selected points to easily obtained data on nonactors, as well as actors. For example, knowledge of the rate of exposure among nonbuyers to an advertisement on television allows us to calculate the total effectiveness, as well as the impact of the advertisement (Zeisel [1947] 1957, pp. 170-173). Not only can demographic or other objective data be used, but also attitudinal and other subjective data from cross-sectional surveys. All that is needed is to ask in a reason analysis some of the same questions that have been asked of a relevant cross section.

### The validity of reason analysis

If all the proper steps of a reason analysis have been followed, can the causes of an action be determined from reasons given by actors? There are two related problems here: the validity of subjective responses and the validity of causal assessment when only actors (or nonactors) are analyzed.

The paramount position of subjective materials in reason analysis may make some researchers uneasy: actors may not know the "real" reasons for their actions, and the researcher may thus be collecting a set of mere rationalizations. Further, there may be additional reasons behind the ones collected, so that the researcher becomes involved in an infinite regression.

If interviews are properly conducted, however, it is the analyst's job to interpret the answers; as Freud (1916-1917) pointed out, whatever a person says must have some meaning. For example, in a study of reasons given by some college graduates as to why they did not choose to go on to graduate school (in the face of the expectation of 80 per cent that they would go on), it was found that lack of motivation was mentioned more often than low grades (Davis 1964). Yet a cross-sectional analysis showed that those who were not going to graduate school had, in fact, lower grades than those who chose to go on. Which finding is

the "real" reason? Did not students with low grades "rationalize" them and say they were not interested? Actually, low grades do lead to low motivation, and low motivation leads to low grades, although a panel study would be necessary to determine which came first [see PANEL STUDIES].

One might say that a "real" reason is simply one that has been built into an accounting scheme. There may be several levels of reasons; for example, people may go to a psychiatrist "because" of an emotional problem that developed at age two or "because" of the influence of a social circle of "friends and supporters of psychotherapy" (Kadushin 1962; 1966). The focus would depend on the purpose of the study.

Finally, there is a good deal of evidence that when the findings of cross-sectional analysis are compared with those of reason analysis, they are found to be similar (Lazarsfeld & Rosenberg 1955, pp. 404-419; Rossi 1955; Burt 1925). Considerable work remains to be done, however, in enhancing and testing the validity of reason analysis, as well as in testing and improving various models of action.

CHARLES KADUSHIN

[See also EVALUATION RESEARCH; EXPERIMENTAL DESIGN, article on QUASI-EXPERIMENTAL DESIGN; INTERVIEWING, article on SOCIAL RESEARCH; SURVEY ANALYSIS, article on METHODS OF SURVEY ANALYSIS.]

### BIBLIOGRAPHY

Many reason analyses are in unpublished commercial research. The best published introductions to the topic are Lazarsfeld & Rosenberg 1955; Lazarsfeld 1959; Zeisel 1947. Especially relevant in the theory of action tradition are Parsons 1937; 1947; 1960; Parsons & Shils 1951; Tollman 1951; Weber 1904-1917. Early but important methodological works in the empirical tradition are Lazarsfeld 1935; 1942. Some empirical works explicitly using accounting schemes and reason analyses are Goode 1956; Kadushin 1958; Katz & Lazarsfeld 1955; Lazarsfeld 1931; Mills et al. 1950; Rossi 1955; Sills 1957.

- BALES, ROBERT F. 1950 *Interaction Process Analysis: A Method for the Study of Small Groups*. Reading, Mass.: Addison-Wesley.
- BARTON, ALLEN H. 1955 The Concept of Property Space in Social Research. Pages 40-54 in Paul F. Lazarsfeld and Morris Rosenberg (editors), *The Language of Social Research*. Glencoe, Ill.: Free Press.
- BERELSON, BERNARD; LAZARSFELD, PAUL F.; and MCPHEE, WILLIAM N. 1954 *Voting: A Study of Opinion Formation in a Presidential Campaign*. Univ. of Chicago Press.
- BURT, CYRIL L. (1925) 1944 *The Young Delinquent*. 4th ed. Univ. of London Press.
- DAVIS, JAMES A. 1964 *Great Aspirations: The Graduate School Plans of America's College Seniors*. Chicago: Aldine.
- FISHER, R. A. (1935) 1960 *The Design of Experiments*. 7th ed. New York: Hafner.

- FREUD, SIGMUND (1916-1917) 1952 *A General Introduction to Psychoanalysis*. Authorized English translation of the rev. ed. by Joan Riviere. Garden City, N.Y.: Doubleday. → First published as *Vorlesungen zur Einführung in die Psychoanalyse*. See especially the second lecture.
- GAUDET, HAZEL 1955 A Model for Assessing Changes in Voting Intention. Pages 428-438 in Paul F. Lazarsfeld and Morris Rosenberg (editors), *The Language of Social Research*. Glencoe, Ill.: Free Press.
- GOODE, WILLIAM J. 1956 *After Divorce*. Glencoe, Ill.: Free Press.
- GURIN, GERALD; VEROFF, JOSEPH; and FELD, SHEILA 1960 *Americans View Their Mental Health: A Nationwide Interview Survey*. Joint Commission on Mental Illness and Health, Monograph Series, No. 4. New York: Basic Books.
- HADDON, WILLIAM; SUCHMAN, EDWARD A.; and KLEIN, DAVID 1964 *Accident Research: Methods and Approaches*. New York: Harper.
- HINKLE, ROSCOE C. 1963 Antecedents of the Action Orientation in American Sociology Before 1935. *American Sociological Review* 28:705-715.
- KADUSHIN, CHARLES 1958 Individual Decisions to Undertake Psychotherapy. *Administrative Science Quarterly* 3:379-411.
- KADUSHIN, CHARLES 1962 Social Distance Between Client and Professional. *American Journal of Sociology* 68:517-531.
- KADUSHIN, CHARLES 1966 The Friends and Supporters of Psychotherapy: On Social Circles in Urban Life. *American Sociological Review* 31:786-802.
- KATZ, ELIHU; and LAZARSFELD, PAUL F. 1955 *Personal Influence: The Part Played by People in the Flow of Mass Communications*. Glencoe, Ill.: Free Press. → A paperback edition was published in 1964.
- KORNHAUSER, ARTHUR; and LAZARSFELD, PAUL F. (1935) 1955 The Analysis of Consumer Actions. Pages 392-404 in Paul F. Lazarsfeld and Morris Rosenberg (editors), *The Language of Social Research*. Glencoe, Ill.: Free Press. → First published by the American Management Association as "The Techniques of Market Research From the Standpoint of a Psychologist."
- LAZARSFELD, PAUL F. (editor) 1931 *Jugend und Beruf*. Jena (Germany): Fischer.
- LAZARSFELD, PAUL F. (1935) 1954 The Art of Asking Why. Pages 675-686 in Society for the Psychological Study of Social Issues, *Public Opinion and Propaganda*. Edited by Daniel Katz et al. New York: Dryden. → First published in Volume 1 of the *National Marketing Review*.
- LAZARSFELD, PAUL F. 1942 The Statistical Analysis of Reasons as Research Operation. *Sociometry* 5:29-47.
- LAZARSFELD, PAUL F. 1959 Reflections on Business. *American Journal of Sociology* 65:1-31.
- LAZARSFELD, PAUL F.; and ROSENBERG, MORRIS (editors) 1955 *The Language of Social Research: A Reader in the Methodology of Social Research*. Glencoe, Ill.: Free Press. → See especially section 5.
- LIPSET, SEYMOUR M.; TROW, MARTIN A.; and COLEMAN, JAMES S. 1956 *Union Democracy: The Internal Politics of the International Typographical Union*. Glencoe, Ill.: Free Press. → A paperback edition was published in 1962 by Doubleday.
- MICHIGAN, UNIVERSITY OF, SURVEY RESEARCH CENTER 1960 *The American Voter*, by Angus Campbell et al. New York: Wiley.
- MILLS, C. WRIGHT; SENIOR, CLARENCE; and GOLDSSEN, ROSE K. 1950 *The Puerto Rican Journey: New York's Newest Migrants*. New York: Harper.
- NAGEL, ERNEST 1961 *The Structure of Science: Problems in the Logic of Scientific Explanation*. New York: Harcourt. → See especially pages 582-588.
- PARSONS, TALCOTT (1937) 1949 *The Structure of Social Action: A Study in Social Theory With Special Reference to a Group of Recent European Writers*. Glencoe, Ill.: Free Press.
- PARSONS, TALCOTT 1947 Introduction. In Max Weber, *The Theory of Social and Economic Organization*. Translated and edited by A. M. Henderson and Talcott Parsons. New York: Free Press.
- PARSONS, TALCOTT 1960 Pattern Variables Revisited: A Response to Robert Dubin. *American Sociological Review* 25:467-483.
- PARSONS, TALCOTT; BALES, R. F.; and SHILS, E. A. 1953 *Working Papers in the Theory of Action*. Glencoe, Ill.: Free Press.
- PARSONS, TALCOTT; and SHILS, EDWARD (editors) 1951 *Toward a General Theory of Action*. Cambridge, Mass.: Harvard Univ. Press. → A paperback edition was published in 1962 by Harper. See especially Chapter 2.
- ROSSI, PETER H. 1955 *Why Families Move: A Study in the Social Psychology of Urban Residential Mobility*. Glencoe, Ill.: Free Press.
- SELVIN, HANAN C. 1957 A Critique of Tests of Significance in Survey Research. *American Sociological Review* 22:519-527.
- SILLS, DAVID L. 1957 *The Volunteers: Means and Ends in a National Organization*. Glencoe, Ill.: Free Press.
- SILLS, DAVID L. 1960 A Sociologist Looks at Motivation. Pages 70-93 in Nathan E. Cohen (editor), *The Citizen Volunteer: His Responsibility, Role, and Opportunity in Modern Society*. New York: Harper.
- SILLS, DAVID L. 1961 On the Art of Asking "Why Not?": Some Problems and Procedures in Studying Acceptance of Family Planning. Pages 26-36 in All India Conference on Family Planning, Fourth, 1961, *Report of the Proceedings: 29th January-3rd February 1961, Hyderabad*. Bombay: Family Planning Association of India.
- SUCHMAN, EDWARD A. 1961 A Conceptual Analysis of the Accident Phenomenon. *Social Problems* 8:241-253.
- TIYAKIAN, EDWARD A. 1965 Existential Phenomenology and the Sociological Tradition. *American Sociological Review* 30:674-688.
- TOLMAN, EDWARD C. 1951 A Psychological Model. Pages 279-361 in Talcott Parsons and Edward Shils (editors), *Toward a General Theory of Action*. Glencoe, Ill.: Free Press.
- WEBER, MAX (1904-1917) 1949 *Max Weber on the Methodology of the Social Sciences*. Translated and edited by Edward Shils and H. A. Finch. Glencoe, Ill.: Free Press.
- ZEISEL, HANS (1947) 1957 *Say It With Figures*. 4th ed., rev. New York: Harper. → See especially Chapters 6-7.
- ZETTERBERG, HANS L. (1954) 1965 *On Theory and Verification in Sociology*. Totowa, N.J.: Bedminster Press.



## REASONING AND LOGIC

Reasoning is thinking enlightened by logic. It can be defined more broadly so that it becomes equivalent to problem solving or critical thinking, but it becomes a special kind of thinking worth separate discussion only when the logical component is stressed. Thus intuitive, imaginative, and trial-and-error thinking are eliminated, no matter how excellent the outcome of such activities may be.

Anyone who contemplates what man's reasoning has accomplished may be moved to curiosity as well as rapture, and there have been many attempts, each in the fashion of its time, to locate the faculty of reason in the scheme of things. Thus to Plato reason was the messenger of the gods. To Augustine reason was the eye of the soul, given to man by God that he may comprehend God. Then, under the impact of the doctrine of evolution, reason, as the highest human capacity, became a critical issue in arguments about the continuity of the species. Although satisfactory methods for tracing the evolution of psychological capacities are lacking (no fossilized thoughts have been uncovered in the chalk beds) and speculations about preliterate man are untrustworthy, reasoning is usually included today with other activities of the organism, such as swallowing and perceiving, that have evolved over the centuries in the struggle for existence.

Logic is a social invention, but it too has a long history of progressive refinement and transmission from one generation to the next, along with such cultural acquisitions as language and clothing. Hence, the reasoning that occurs today can be considered an adaptive natural function operating with cultural products that have been developed to fit the minds that use them.

Logic apparently developed, in India as in Greece (Bocheński 1956), out of techniques for refuting arguments. Certainly it was devised by people for people. Thus it is not surprising that many philosophical psychologists of the nineteenth century treated logic as the laws of thought, as idealized statements of how people reason. Today most logicians assert that they are concerned with formal relations between propositions but not with the thinking processes that produce them. Psychology is concerned with thinking, including thinking enlightened by logic, as a fallible human activity. The empirical study of such logical thinking began in England early in the twentieth century (see Wheeler 1958) when psychologists and educators used syllogisms for tests of ability and

as materials for the investigation and training of reasoning. Since then psychological research has been directed toward traditional more than modern logic and deductive more than inductive reasoning, although some attention is given to inductive reasoning in discussions of concept formation.

**Criteria of reasoning.** Since many logical expressions were taken originally from the resources of the general language, the verbal patterns of logical statements often resemble the verbal patterns of ordinary discourse. The famous syllogism "All men are mortal; Socrates is a man. Therefore Socrates is mortal" looks something like a sequence of sentences that can be read by anyone with ordinary reading ability. But a logical system also includes technical concepts that are not in the general vocabulary, rules of combination that are not rules of grammar, and a metalanguage for making statements about statements. These logical skills depend on reading habits and abstract thinking, but they are special formal achievements. When someone uses variables to stand for propositions and writes, "If P, then Q," or rearranges the verbal patterns of ordinary discourse into a standard logical form, he has gone beyond the general language. When one uses these concepts and procedures to test an argument for inconsistency or to generate nonobvious inferences from a problematic situation, his thinking can be described as reasoning.

**Difficulties in reasoning.** Reasoning, as defined above, may occur during the solution of many problems, but it occurs most obviously during the solution of problems of formal logic—for example, syllogisms. Hence logic supplies materials for the psychological investigation of reasoning, as chemistry supplies materials for the investigation of smell. Such reasoning problems are difficult for the ordinary person, who solves most of his daily problems adequately. The laws of thought may be related, as ideals, to actual thinking, but the relation is tenuous; man does not often reach his ideals. Deductive reasoning especially is a cultivated performance, a kind of game played by rules that are not quite like those of life played in earnest. Correct solution of a logic problem requires a logical approach, a single-minded restriction of attention to the implications of sentences, although the habits acquired in years of reading and listening direct attention to connotations and other verbal and emotional associations. Children and intelligent adults without training in logic make many errors in working reasoning problems, and research on the nature of the errors illuminates the reasoning process itself.

*Terminological difficulties.* The terms of a logical statement may cause difficulty; syllogisms written in familiar terms are somewhat easier than the same syllogisms written in letters or unfamiliar terms (Wilkins 1928). The quantifier "some" is a special difficulty because many people include with it the notion of "but not all." This is a reasonable extension of ordinary reading habits and is allowed in Indian logic (Bocheński [1956] 1961, p. 437) but leads to technical errors in the usual systems of Western logic.

*Abstractness of materials.* Obviously some of the difficulty in reasoning is due to the abstract nature of the materials. When children have to learn concepts and principles of different levels of abstraction and to use them according to instructions, more errors are made on the more abstract problems (Long & Welch 1942; Welch & Long 1940). As objects of thought, to be retained and processed in immediate memory despite interference from other objects of thought, classes are more difficult than perceived objects, and implications of propositions are more elusive than perceived relations.

*Order of presentation.* The order in which logical statements appear can also cause trouble because when analysis of the argument requires rearrangement of the statements, certain common habits of reading and thinking may interfere. Consider the following:

John is taller than Mary;  
Mary is taller than David.  
Is John taller than David?

This is an easy problem because the order in which the thinker gets the information is a convenient order for him to process it. Now consider

Mary is shorter than John;  
Mary is taller than David.  
Is John taller than David?

Problems of this second type of logical structure are more difficult, whether difficulty is measured in time required for correct conclusion (Hunter 1957) or in number of errors (De Soto et al. 1965). Unlike the first problem, which is so straightforward that the thinker can almost read the conclusion directly from the premises, the second has a more complex structure, and the relations have to be somehow rearranged before the logical implications become apparent. Many people do this by drawing little charts on paper, or by making gestures with their hands, or by some other kind of imaginary spatial representation. Thus problems stated in a relation that is easily visualized, such as "taller than," are easier for English school children at age 11 than problems stated in terms of the

relationship "warmer than" or "happier than." This difference in difficulty is not found at age 16, when abstract skills presumably have replaced visualization (Hunter 1957).

*Direction of relations.* Even when the order of the argument is straightforward, one direction may be easier for solution than the other. American college students can read "A is to the left of B; B is to the left of C" and follow it more readily to a conclusion than the logically equivalent "C is to the right of B; B is to the right of A." Similar problems based on the relation "better than" are easier than those based on "worse than," and those stated in terms of the relation "above" are easier than those stated in terms of "below." Habits of reading and writing make it relatively easy to represent "left" on the left and proceed rightward, or "better" or "above" or "taller" on top and proceed downward. The reverse procedure can be followed, but it is less habitual and more subject to error. Propositions stated in "lighter than" and "darker than" are equally difficult, however, because one can be represented spatially as easily as the other (De Soto et al. 1965).

*Mixing direction of relations.* The difficulty is increased when one premise is stated in one direction—for example, "shorter than"—while the other is stated in the other direction—"longer than." Even though the proposition "Mary is shorter than John" is equivalent to "John is taller than Mary," the first must be rearranged in the second form if it is to be aligned with "Mary is taller than David." This rearrangement takes time and attention, and if the thinker has not memorized the rules he may make mistakes. Thus whenever such rearrangement is required by the structure of the premise combinations, time and errors increase (De Soto et al. 1965; Hunter 1957).

*Location of the middle term.* If C and A are related by their relation to B, many thinkers construct a three-term structure, either verbally or spatially, with the distributed term, B, in the middle. This is easier when the first term in a premise is an end of the structure to be attained rather than the middle. Hence, it is easy to relate A and C in these two cases: "A is better than B; C is worse than B." "C is worse than B; A is better than B." And in these two it is hard: "B is worse than A; B is better than C." "B is better than C; B is worse than A" (De Soto et al. 1965).

*Conversion and atmosphere errors.* Research on arguments that involve categorical propositions of the form "All X is Y" has uncovered many errors due to invalid conversion. The converse of the proposition "All men are rational" would prop-



erly be a statement such as "The class of rational beings includes man." However, when a sample of Frenchmen was asked to convert this proposition, only a few did it properly; many more wrote statements such as "Rationality is the property of man." The conversions chosen for various propositions depended as much on the meaning of the proposition as on its logical form (Oléron 1964).

"All  $X$  is  $Y$ " is frequently converted to "All  $Y$  is  $X$ " because ordinary reading habits are carried over to a logical proposition. When the proposition is written  $X \subset Y$  and read "All  $X$  is included in the class  $Y$ ," invalid conversion is less common. Consider the syllogism

All  $S$  is  $M$ ;

All  $P$  is  $M$ .

Therefore, all  $S$  is  $P$ .

One hypothesis for the common acceptance of such an invalid argument is the "atmosphere effect." The untrained reader who does not know the rules and cannot follow through the abstract implications seizes on nonlogical verbal cues. The word "all" in the premises suggests a universal affirmative atmosphere, so when he sees the same word in the conclusion he accepts it. The word "some" in the premises wafts a favorable atmosphere around a conclusion stated in "some." This hypothesis about atmosphere effect predicts that certain invalid arguments will be accepted more often than others, and when Sells (1936) presented syllogisms written in terms of  $X$ ,  $Y$ , and  $Z$  to college students, the results agreed with the predictions fairly well. Similar research using syllogisms written in familiar words as well as letters, with the conclusion to be chosen from five alternative conclusions, also confirmed most of the predictions from atmosphere effect (Morgan & Morton 1944).

Another interpretation of the results of such research emphasizes invalid conversion strengthened by the habit of probabilistic inference. Many statements may be converted simply, like arithmetic statements of the form "All 90° angles are right angles"; hence, acceptance of a conversion is often reinforced in ordinary problem solving, and one who has not been warned will transfer these intermittently reinforced habits to the syllogism game. For example, if the second premise in the above syllogism is converted from "All  $P$  is  $M$ " to "All  $M$  is  $P$ ," the conclusion is readily accepted as valid. This interpretation accounts for acceptance of invalid arguments approximately as well as predictions from atmosphere effect [see Chapman & Chapman 1959; see also PROBABILITY; and the biography of BRUNSWIK].

Although the atmosphere effect and the tend-

ency toward invalid conversion were intended as alternative interpretations of the most common errors in syllogistic reasoning, recent evidence indicates that they are independent (Simpson & Johnson 1966). Those college students who make one type of error do not, as a rule, make the other type. Furthermore, differential training aimed specifically at the atmosphere effect reduced the atmosphere error but not the conversion error. Training aimed specifically at the conversion error tended to have a parallel effect on conversion errors.

*Personal biases.* Since a logical course through an argument to a conclusion is often hard to follow, the distracting influence of bias is easily demonstrated. Syllogisms written in controversial terms, such as "communists" and "war," are generally more difficult to evaluate than syllogisms of the same form written in neutral terms, such as "whales" and "water" (Lefford 1946). The effects of bias on acceptance of specific conclusions has also been demonstrated by variation in the terms of the arguments. Graduate students with no training in logic judged syllogisms with respect to the validity of the arguments and later were asked if they personally agreed with the conclusions. Most of their errors were due to acceptance of conclusions they liked and rejection of conclusions they disliked (Janis & Frick 1943).

*Cognitive theories.* Current interest in cognitive theories among social scientists has encouraged more formal models of the interaction between logical reasoning and personal desires in the expression of attitudes and beliefs. One of these, the "logic model" of McGuire (1960), begins by asking people to estimate the probabilities of premises and conclusions, then studies whether these are related as the model predicts, and examines the discrepancies from the predictions. Probabilistic reasoning, according to this model, requires that the probability of a conclusion be quantitatively related to the combination of the probabilities of the premises. Research with high school and college students disclosed a moderate relation; even on controversial issues these people were somewhat logical. A moderate relation was also found between acceptance of conclusions and ratings of desirability of premises—which can be interpreted as wishful thinking (influence of desires on beliefs) or rationalization (influence of beliefs on desires), or both.

A more subtle logical influence was demonstrated by McGuire when new information increasing the probabilities of minor premises resulted in higher estimates of probabilities of logically related but not explicitly mentioned conclusions.

However, this effect did not last long. Apparently beliefs represent a balance between desire and reason; new information may force logical re-examination of beliefs and thus temporarily increase logical consistency, but away from the influence of the new information desires weigh heavier in the balance, and logical consistency suffers. [See THINKING, *article on COGNITIVE ORGANIZATION AND PROCESSES.*]

In these studies the psychologist has taken a system of logic as a model and has tried to describe how the thinking of ordinary people deviates from this model. But if they deviate so often in so many ways, why use this model to describe their behavior? The answer would be that no other model yet proposed describes the facts with fewer deviations. Ordinary people apparently do strive, with varying effort and skill, toward logical consistency as a goal, they construct visual aids to guide them, and they come closer to the goal with practice. The errors described above are due in general to (1) differences between the habitual intellectual activities of untrained people and the technical procedures of logical systems, (2) overloading of immediate memory by the necessity for rearranging and recoding abstract materials, and (3) bias for or against the conclusions.

**Mathematical reasoning.** Among logicians the prevalent view is that all inference, whether in chemistry, mathematics, or elsewhere, is logical. Among educators, however, mathematical reasoning appears to be a special practical skill. It is identified as such and taught to school children much more seriously than reasoning in other areas, such as social science. Errors of mathematical reasoning, analogous to the errors listed above, are exposed routinely. Precise statement is emphasized. Thus the intelligent adult who makes glaring errors in syllogistic reasoning, prompting the question whether man is naturally logical, has had years of experience in mathematical manipulations and does not make comparable mathematical blunders. The dependence of mathematical reasoning on mathematical training is obvious, so the question whether man is naturally capable of mathematical reasoning does not arise.

**Sequences of reasoning operations.** The time spent on the solution of problems has been used as a measure of reasoning difficulty, and some problems can be divided so that the time devoted to component operations can be recorded in sequence. Conventionally it is assumed that the solution of an analogy problem of the form "Feline is to canine as cat is to . . . ?" begins with induction of the relation between the first two terms

and proceeds to application of this relation to the third term for deduction of the solution. This assumption can be tested with the aid of suitable exposure apparatus that separates the time the thinker spends studying the first pair of words, called induction time, from the time he spends studying the third word and writing a solution word, called deduction time. Problems in which the first pair of words is difficult, as above, require relatively long induction time, and those in which the second pair is difficult—such as "Lose is to win as liability is to . . . ?"—require relatively long deduction time (Johnson 1962). The two logical operations may overlap in time, but the obtained time differentials show that the overlap cannot be large.

Sequences of operations in solving problems of proof in modern logic, specifically in the sentential calculus of Alfred North Whitehead and Bertrand Russell, have been charted by requiring the thinker to state the rule for each operation as he performs it (Anderson 1957). Such research would hardly be worthwhile if reasoning always progressed deductively, in a straight line from premises to conclusions, but in these problems, as in others, many thinkers plot their course by the destination as well as by the starting point. In problems that required six steps for proof the first steps were relatively easy because the implications from the data were clear, and the last steps were relatively easy because the logical requirements for reaching the goal became apparent. But in the middle the sequence of steps was not so obvious, and irrelevant operations occurred frequently.

Computer programs also have been written for discovering proofs in the sentential calculus, and comparison of the computer's behavior—in respect to intermediate steps, failures, and effects of sequences of problems—with people's behavior may lead to a better understanding of human reasoning.

**Learning to reason.** A sound mind in a sound body is not drawn irresistibly to sound conclusions. Reasoning is a cultivated performance that depends on practice. A college course in logic improves skill in handling logic problems, and Navy enlisted men have acquired some facility in the sentential calculus in two hours of special instruction (Moore & Anderson 1954). Special instructions with diagrams and examples helped a group of adults to reach a high degree of accuracy and to demonstrate that the distorting effects of bias can be minimized (Henle & Michael 1956). When sequences of related reasoning problems were arranged so that a fallacious inference would contradict a preceding inference, college students



learned to detect their own errors and made definite improvement (Wason 1964). Such improvement would not occur if formal reasoning were a routine performance that is practiced daily.

From the psychological point of view, training in logical reasoning departs from the general language and consists of learning special concepts, procedures, purposes, much as does training in other systematic disciplines. Practice in putting verbal patterns in a class—for instance, primitive constant—and naming them increases skill in dealing with the class as a unit. Rules for operations, such as conversion, are memorized and practiced. Ordinary discourse is analyzed and the arguments rearranged into formal patterns. The purposes and restrictions of formal logic are distinguished from those of ordinary communication.

Many of the difficulties listed above are reduced by techniques that make abstract relations more concrete, such as Euler's circles, Venn diagrams, and truth tables. These techniques may be considered recoding techniques in that the information which appears as sentences, numerals, or shapes is transformed into some other kind of information that is more easily processed and matched to formal inference patterns. Psychologists have not yet worked out the principles that govern this high-level cognitive learning, except for the learning of concepts, and the terminology of the logician may not coincide with psychological processes, but the intuitions of interested teachers have been helpful to their students. Some people learn these skills by themselves, and some make better use of what they have been taught than others. Hence, when intelligent adults who have not studied logic are compared on reasoning tests with those who have, although the average scores are different, there is considerable overlap between the groups [see Morgan & Morgan 1953; see also CONCEPT FORMATION].

*Transfer of training.* Improvement on reasoning problems as a result of special practice is rather specific; the transfer to other problems is not large (see Johnson 1955). As logic continues to develop, its technical language will probably diverge even more than now from the general language, and the amount of transfer may be even less. Whereas the older logicians analyzed the inconsistencies in ordinary language, the modern logicians are creating formal languages for their own purposes which, though they include some familiar constants—"if," "only," "not," "or"—are not adapted to the contexts and purposes of ordinary communication. Nevertheless, any school subject can be taught with emphasis on induction

of generalizations, precise statements of relationships, organization of data into cognitive structures, logical inferences from these structures, separation of inferences from observations, and detection of common fallacies. Some teachers have tried to do this in the past, and post-Sputnik trends may reinforce the attempt.

*Inference of reasoning in problem solving.* When the thinker is faced by a logic problem of appropriate difficulty, he will try to use whatever logical patterns and procedures he has learned, and his performance will improve with training in logic; hence the psychologist may say that reasoning is a substantial part of his problem-solving effort. And it is a tenable hypothesis that reasoning also occurs during the solution of problems that are not logic problems per se, but the confirmation of this hypothesis is complicated. Throughout the history of this topic, problem solving that has been interpreted as reasoning has also been interpreted in terms of instinct or trial-and-error learning, or some alternative hypothesis.

Correct solution of a problem is not itself sufficient evidence that reasoning has occurred, nor, as Henle (1962) has argued, is failure good evidence for the absence of reasoning. But there are at least two more specific types of evidence. One is the demonstration that the solution is mediated by a specific logical structure, such as a syllogism. The hypothesis that a problem is solved through the mediation of a syllogism is supported when the thinker produces verbal comments or diagrams along with solutions, all of which can be better predicted from the syllogism hypothesis than from any alternative hypothesis. The other type of evidence is improvement in problem solving as a result of training in logic, which involves one of the standard transfer designs of experimental psychology. If training in logical reasoning improves performance in problem solving as compared to initial performance, the final problem solving must include something, such as logical forms and operations, carried over from the logical training. Scattered results to date indicate that the transfer is small and variable, depending on the nature of the training, the nature of the problems, and the abilities of the individuals.

*Reasoning abilities.* It is relatively easy to construct a reliable test of reasoning ability, simply by assembling a number of logic problems, such as syllogisms. This was done a half century ago in England (Burt 1919); since then reasoning has been considered an important component of general intelligence. People who score high on such tests of reasoning ability, with noncontroversial

conclusions, are able to discount their own biases fairly well when judging the validity of arguments with controversial conclusions (Feather 1964). Tests of arithmetic reasoning ability can also easily be assembled, and this ability has long been distinguished from facility in simple arithmetic computations. When such reasoning tests and general intelligence tests are given to the same people, the scores are correlated because the abilities required for success on each have much in common. There are several plausible interpretations for the correlations obtained between pairs of tests, and many statistical analyses have been directed toward one or more of the following: (1) Reasoning is the central factor in intelligence. (2) Reasoning is one factor among several. (3) Reasoning itself can be split into several factors.

In the most sophisticated attack on this problem (Guilford 1959; Merrifield 1966) some abilities are identified according to the materials that are processed: figural, symbolic, and semantic. Others are identified according to the products that result: units, classes, relations, systems, transformations, and implications. In this formulation reasoning abilities could include classificatory, relational, and systemic induction as well as relational and implicational deduction.

DONALD M. JOHNSON

[Directly related are the entries CONCEPT FORMATION; PROBLEM SOLVING; THINKING; SIMULATION, article on INDIVIDUAL BEHAVIOR. Other relevant material may be found in ATTITUDES; DECISION MAKING; DECISION THEORY; DEVELOPMENTAL PSYCHOLOGY, article on A THEORY OF DEVELOPMENT; INTELLIGENCE AND INTELLIGENCE TESTING; MATHEMATICS.]

#### BIBLIOGRAPHY

See Strawson 1952 for clarification of the relations between general language and the language of logic. See Kaplan 1964 for a discussion of "logic in use," especially for the social scientist, as distinguished from a more formal or "reconstructed" logic. For a somewhat different analysis of reasoning abilities from the British point of view, see Wheeler 1958. For a French review, see Oléron 1957. For an analysis of reasoning abilities in a Philippine sample, see Guthrie 1963. For information on attainment of logical symbols by children of different ages, including deaf and retarded children, see Furth & Youniss 1965 and Youniss & Furth 1964.

ANDERSON, SCARVIA B. 1957 Problem Solving in Multiple-goal Situations. *Journal of Experimental Psychology* 54:297-303.

BOCHEŃSKI, I. M. (1956) 1961 *A History of Formal Logic*. Translated and edited by Ivo Thomas. Univ. of Notre Dame Press. → First published as *Formale Logik*.

BURT, CYRIL 1919 The Development of Reasoning in School Children. *Journal of Experimental Pedagogy* 5:68-77, 121-127.

CHAPMAN, LOREN J.; and CHAPMAN, JEAN P. 1959 Atmosphere Effect Re-examined. *Journal of Experimental Psychology* 58:220-226.

DE SOTO, CLINTON; LONDON, MARVIN; and HANDEL, STEPHEN 1965 Social Reasoning and Spatial Paralogic. *Journal of Personality and Social Psychology* 2:513-521.

FEATHER, N. T. 1964 Acceptance and Rejection of Arguments in Relation to Attitude Strength, Critical Ability, and Intolerance of Inconsistency. *Journal of Abnormal and Social Psychology* 69:127-136.

FURTH, HANS G.; and YOUNISS, JAMES 1965 The Influence of Language and Experience on Discovery and Use of Logical Symbols. *British Journal of Psychology* 56:381-390.

GUILFORD, J. P. 1959 Three Faces of Intellect. *American Psychologist* 14:469-479.

GUTHRIE, GEORGE M. 1963 Structure of Abilities in a Non-western Culture. *Journal of Educational Psychology* 54:94-103.

HENLE, MARY 1962 On the Relation Between Logic and Thinking. *Psychological Review* 69:366-378.

HENLE, MARY; and MICHAEL, MIRIAM 1956 The Influence of Attitudes on Syllogistic Reasoning. *Journal of Social Psychology* 44:115-127.

HUNTER, IAN M.-L. 1957 The Solving of Three-term Series Problems. *British Journal of Psychology* 48:286-298.

JANIS, IRVING L.; and FRICK, FREDERICK 1943 The Relationship Between Attitudes Towards Conclusions and Errors in Judging Logical Validity of Syllogisms. *Journal of Experimental Psychology* 33:73-77.

JOHNSON, DONALD McEWEN 1955 *The Psychology of Thought and Judgment*. New York: Harper.

JOHNSON, DONALD McEWEN 1962 Serial Analysis of Verbal Analogy Problems. *Journal of Educational Psychology* 53:86-88.

KAPLAN, ABRAHAM 1964 *The Conduct of Inquiry: Methodology for Behavioral Science*. San Francisco: Chandler.

LEFFORD, ARTHUR 1946 The Influence of Emotional Subject Matter on Logical Reasoning. *Journal of General Psychology* 34:127-151.

LONG, LOUIS; and WELCH, LIVINGSTON 1942 Influence of Levels of Abstractness on Reasoning Ability. *Journal of Psychology* 13:41-59.

MCQUIRE, WILLIAM J. 1960 A Syllogistic Analysis of Cognitive Relationships. Pages 65-111 in *Attitude Organization and Change: An Analysis of Consistency Among Attitude Components*, by Milton J. Rosenberg et al. Yale Studies in Attitude and Communication, Vol. 3. New Haven: Yale Univ. Press.

MERRIFIELD, PHILIP R. 1966 An Analysis of Concepts From the Point of View of the Structure of Intellect. Pages 19-33 in Herbert J. Klausmeier and Chester W. Harris (editors), *Analyses of Concept Learning*. New York: Academic Press.

MOORE, OMAR K.; and ANDERSON, SCARVIA B. 1954 Modern Logic and Tasks for Experiments on Problem Solving Behavior. *Journal of Psychology* 38:151-160.

MORGAN, JOHN J. B.; and MORTON, JAMES T. 1944 The Distortion of Syllogistic Reasoning Produced by Personal Convictions. *Journal of Social Psychology* 20:39-59.

MORGAN, WILLIAM J.; and MORGAN, ANTONIA B. 1953 Logical Reasoning: With and Without Training. *Journal of Applied Psychology* 37:399-401.



- OLÉRON, PIERRE 1957 *Les composantes de l'intelligence d'après les recherches factorielles*. Paris: Presses Universitaires de France.
- OLÉRON, PIERRE 1964 *Les activités intellectuelles*. Paris: Presses Universitaires de France.
- SELLE, SAUL B. 1936 The Atmosphere Effect: An Experimental Study of Reasoning. *Archives of Psychology* No. 200. → The entire issue is devoted to reasoning studies.
- SIMPSON, MILES E.; and JOHNSON, DONALD M. 1966 Atmosphere and Conversion Errors in Syllogistic Reasoning. *Journal of Experimental Psychology* 72:197-200.
- STRAWSON, P. F. 1952 *Introduction to Logical Theory*. London: Methuen; New York: Wiley.
- WASON, P. C. 1964 The Effect of Self-contradiction on Fallacious Reasoning. *Quarterly Journal of Experimental Psychology* 16:30-34.
- WELCH, LIVINGSTON; and LONG, LOUIS 1940 The Higher Structural Phases of Concept Formation of Children. *Journal of Psychology* 9:59-95.
- WHEELER, DOROTHY 1958 Studies in the Development of Reasoning in School Children. *British Journal of Statistical Psychology* 11:137-159.
- WILKINS, MINNA C. 1928 The Effect of Changed Material on Ability to Do Formal Syllogistic Reasoning. *Archives of Psychology* No. 102. → The entire issue is devoted to reasoning studies.
- YOUNISS, JAMES; and FURTH, HANS G. 1964 Attainment and Transfer of Logical Connectives in Children. *Journal of Educational Psychology* 55:357-361.

## RECIDIVISM

See PENOLOGY, article on PROBATION AND PAROLE.

## RECIPROCITY

See INTERACTION, article on SOCIAL EXCHANGE.

## RECREATION

See COMMUNICATION, MASS, especially the article on TELEVISION AND ITS PLACE IN MASS CULTURE; DRAMA; FILM; LEISURE; LITERATURE; PLANNING, SOCIAL, article on REGIONAL AND URBAN PLANNING.

## REDFIELD, ROBERT

Robert Redfield (1897-1958) was born in Chicago, Illinois. His father, after whom he was named, was a successful corporation lawyer, and his mother, Bertha Dreier Redfield, was the daughter of the Danish consul in Chicago. Both his mother's and his father's families were well-knit groups and included people of education and accomplishment. His father's relatives were early Americans, some of whom had moved to Illinois in the 1830s and acquired land there. Redfield spent the summers and holidays of his youth in the country, where he acquired a deep sense of the social changes that in

a few generations had transformed a wilderness landscape, where pioneers encountered Indians, into an area of commercial farmlands adjoining a great industrial city. In later life he became the first and, for many years, the leading anthropologist to direct attention to the processes of social and cultural change that characterize the relationships between folk and urban societies.

His career, from the publication of his first paper in 1926 until his death in 1958, spanned a period of rapid growth in anthropological research and in the theoretical maturation of the social sciences. The University of Chicago was a center for these developments, and Redfield's life and work were intimately linked with that institution. Redfield's father-in-law, Robert Park, an influential Chicago sociologist, helped to give his career its direction.

The first department of sociology in the United States had been founded at the University of Chicago in 1892, and in the period before World War I, W. I. Thomas, John Dewey, Thorstein Veblen, George Herbert Mead, and other prominent innovators in the social sciences were on the faculty of the university. Park joined the faculty in 1915 and soon began the program of field research in urban life that became the hallmark of the Chicago school of sociology in the 1920s and the 1930s. The Chicago sociologists, largely because of Park's influence, investigated at firsthand the kinds of people who inhabited the city and the ecological and social processes that shaped their lives. They saw Chicago as a mosaic of different ways of life, which they recorded in the same way that functionalist anthropologists were then beginning to record the ways of life in exotic cultures. When Redfield began graduate study in anthropology in 1924, instruction was given in the sociology department; a separate department of anthropology was not created until 1930.

Redfield did not easily arrive at the decision to become an anthropologist. He had been educated at the University of Chicago, graduating from its high school in 1915 and receiving an A.B. in 1920 and a degree from the law school in 1921. He had left college in 1917 to drive an ambulance with the French Army on the western front. After World War I he studied biology at Harvard University, but, unsatisfied by that subject, he was soon persuaded to return to Chicago and study law. In 1920 he married Margaret Lucy Park and the following year, on receiving his law degree, began practice with his father's law firm.

Discontented with law practice and uncertain about what he wanted to do, Redfield visited Mex-

ico in 1923. There he and his wife caught the excitement of the revival of Mexican culture that followed the revolution, and they met and were impressed by the work of Manuel Gamio, an ethnologist and social reformer. Returning to Chicago, Redfield was encouraged by Park to take up the study of anthropology. He did so and returned to Mexico in 1926 to conduct field research in Tepoztlán, an Aztec community near Mexico City.

Redfield's study of Tepoztlán, published in 1930, was in some ways characteristic of other functionalist studies of the period, which were stimulated by Bronislaw Malinowski's work in the Trobriand Islands and A. R. Radcliffe-Brown's study in the Andaman Islands. However, while most anthropologists of a functional persuasion were studying primitive societies that had remained isolated from the influence of civilization and went to work alone on remote islands or among interior tribes, Redfield, accompanied by his wife and two small children, conducted field research in a peasant community that for many centuries had been part of a civilization.

*Tepoztlán* is a modest book, attempting little more than a sympathetic and holistic portrayal of the community, but in writing it Redfield was aware of larger issues. He saw that Tepoztlán resembled peasant villages in other civilized areas and was attentive to the changes that these communities underwent in contact with urban industrial civilization. In 1930, the year *Tepoztlán* was published, he was appointed a research associate of the Carnegie Institution of Washington, and over the next 16 years he conducted field research in Yucatan and Guatemala in connection with that institution. In this work, following a suggestion of Park's, he developed an ideal-type construct of folk society as a way of making more systematic analyses of the transition within civilizations from folk to urban communities.

Meanwhile, he had joined the faculty of the University of Chicago in 1927 and had received his Ph.D. degree from the university in 1928. His talents commanded early recognition, and he rose rapidly in the academic hierarchy, becoming a professor of anthropology and dean of the social science division in 1934.

Redfield's work in Yucatan in the 1930s was an ambitious controlled comparison of four communities—Mérida, the capital city; a provincial railroad town; a peasant village; and a tribal community in the forests of Quintana Roo. He was aided in this research by his wife and by other fieldworkers, particularly Alfonso Villa Rojas, a young schoolteacher whom he met in Yucatan and trained

both in the field and at the University of Chicago and who subsequently became one of Mexico's leading anthropologists.

There were several aspects to this research. An ethnographic record of contemporary Mayan culture was made, one result of which was *Chan Kom: A Maya Village* (1934), written in collaboration with Villa Rojas. Also, since the regional culture of Yucatan was a centuries-old amalgam of Spanish and Mayan traditions, the project touched upon problems of acculturation, which was a new subject for anthropology in the 1930s. But Redfield subordinated both of these historical aspects of the work to the study of more abstract and, in his view, more scientific problems of social change. For Redfield saw history and science as

... opposing methods: the scholarly pursuit of special knowledge of particular fact, on the one hand, and such pursuit of special knowledge of general fact, on the other hand. The historian, in this sense, is concerned with the uniqueness of his data; each proposition in which he expresses his facts has reference to particular space and time. . . . The scientist . . . finds singular propositions referring to definite place and time of service only as they illustrate or contribute to general propositions as to the nature of classes of phenomena. (1962-1963, vol. 1, pp. 12-13)

In *The Folk Culture of Yucatan* (1941) Redfield used the sociological variables of relative size, isolation, and homogeneity to construct a hypothetical continuum between folk and urban societies. The city, town, village, and tribal communities that he and his co-workers chose to study were seen to occupy different positions along this continuum. A synchronic comparison of variations between them revealed causal relationships between the progressive degrees of cultural disorganization, secularization, and individualization. Redfield made no claim to great theoretical innovation but based his thought upon ideas that went back to Henry Maine's contrast between a contractual society and one of status, Ferdinand Tönnies' contrast between *Gemeinschaft* and *Gesellschaft*, and similar concepts in the works of Émile Durkheim, Howard Becker, and other sociologists (1962-1963, vol. 1, pp. 232-234). In this respect, the considerable originality of his work in Yucatan was analogous to that of contemporary biologists who gave new precision to the concept of natural selection by observing and analyzing in nature and in the laboratory actual evolutionary transformations within species.

Outside the circle of his students and colleagues at the University of Chicago and a small group of Mexican anthropologists, the reception of *The Folk*



*Culture of Yucatan* was disappointing. Other anthropologists admired the scale and elegance of the research but were indifferent to Redfield's interests or failed to grasp his points. The most influential work of the period was being done by British ethnologists who were inventing ways to describe the stable social structures of tribal societies and, among American anthropologists, by configuration-alists who were studying the ramifications of cultural integration. In contrast, Redfield's subject was social change and cultural disorganization. In addition, his method of analysis was unfamiliar to anthropologists. The heuristic use of a model folk society was misunderstood by historically minded ethnologists, who saw discrepancies between the model and particular folk communities as flaws in the model rather than as invitations to analyze the causes of social variations. Social scientists who knew the works of Park, Louis Wirth, and other Chicago sociologists responded to Redfield's work from the beginning, and as the generation of social anthropologists trained after World War II turned in greater numbers toward the study of peasant communities and the regional structures and networks of civilizations, Redfield's ideas became increasingly influential.

In 1948 Redfield returned to Chan Kom, this time to observe the changes that had occurred since his work there in the early 1930s. *A Village That Chose Progress* (1950) is a narrative history of the community that added to the earlier comparative analysis an existential appreciation of the ways in which the villagers' expectations, moods, and purposes had influenced the course of social change.

From the time of his restudy of Chan Kom until his death ten years later, Redfield continued to teach at the University of Chicago, becoming the Robert Maynard Hutchins distinguished service professor in 1953. He also traveled extensively, lecturing at colleges and universities in China, India, Europe, Puerto Rico, and the United States, and, with the aid of a large grant from the Ford Foundation, launched a program of seminars and research projects that brought anthropologists and humanists together for the comparative study of civilizations. Much of the work of these seminars and projects was published in *The Primitive World and Its Transformations* (1953), *The Little Community* (1955), *Peasant Society and Culture* (1956), and in articles and lectures, some of which were reprinted in a two-volume collection of his papers (1962-1963).

In these last years Redfield worked with Milton Singer and other scholars, defining and broadening the scope of anthropological studies of civilization.

Thinking of civilizations as members of a class, "a formed thing of the mind," he rejected boxlike definitions and preferred "a form of thought that sees a society to be civilized 'to the extent and in the respects that' it has one or more of very many qualities" (1962-1963, vol. 1, p. 370). The continuum between folk and urban societies, which he had earlier defined by constructing a model folk society, he now approached from the opposite end, replacing the term "urban society" with the more comprehensive "civilized society."

He conceived of civilized societies as historical structures with characteristic life styles and modes of relationship. Urban communities are one dimension of such societies; the structures of rural and peasant communities are another. Drawing upon his own work in Middle America and upon the research of other social anthropologists, he explored the usefulness of the concepts of social fields, centers, networks, and levels to describe the articulation of peasant communities with the larger societies of which they are parts.

He also conceived of civilizations in cultural terms as systems of coexisting and interdependent "Great and Little Traditions," the former being part of the idea systems—the science, philosophy, and fine arts—of the critical and reflective elite and the latter being part of the folk arts, lore, and religion of the common people. With Singer, a trained philosopher who became an anthropologist and a student of India, he analyzed different types of cities and processes of urbanization, in particular distinguishing orthogenetic cities, in which the moral order of the countryside is elaborated into the Great Traditions of indigenous civilizations, and heterogenetic cities, "where local cultures are disintegrated and new integrations of mind and society are developed" (1962-1963, vol. 1, p. 334).

These and other concepts for the comparative study of civilization were developed by Redfield within an evolutionary conception of cultural history. Thus, orthogenetic cities are centers where earlier stages in the development of Great Traditions could be studied, while the later and more advanced stages are centered in heterogenetic cities. Furthermore, in the long span of human history, Redfield saw the emergence of civilizations out of the folk societies of paleolithic times as a multilineal series of events in cultural evolution, showing progressive trends in the moral life of humanity as well as in technology and the division of labor (1953).

When Redfield took up the holistic study of civilization, he found a common ground for the exchange of ideas between social scientists and

humanists, and he entered this exchange with enthusiasm. As dean of the social science division of the University of Chicago from 1934 to 1946, he had worked to build a community of scholars free from the parochialism of academic disciplines. In the 1950s, at the height of his career as an anthropologist, he lectured and wrote on the nature and methodology of his own discipline, emphasizing its relation with the humanities (1962-1963, vol. 1, pp. 33-139). He also initiated a series of conferences in which students of Chinese, Indian, and Islamic civilizations—both humanists and anthropologists—considered common problems. Throughout his career he also lectured and wrote occasionally on the humane goals of education and the social sciences (1962-1963, vol. 2, *passim*), and this, together with his other work, gained for him a reputation as a humanist and a social philosopher.

Philosophically, Redfield was a naive realist, that is, he assumed a reality independent of the knower of it and treated ideas about the world as forms of thought that were more or less adequate according to the uses made of them. This pragmatic and contextual approach to knowledge stood in contrast to the positivism of social scientists who looked primarily to physics for a model of the sciences and who thought Redfield's approach was more "humanistic" than it actually was—humanistic in the sense of romantic, imprecise, unscientific, or even antiscientific. But Redfield's humanism was of a different order. A strong advocate of freedom of thought and expression, he believed in "the creative values of human life in its social process, the competence of reason and scientific method to aid in the understanding of human actions, . . . and the responsibility of man to man in the affairs of life" (1962-1963, vol. 1, p. ix). This form of humanism, essentially an openness to life, was a fundamental aspect of Redfield's scientific outlook. He claimed that the ultimate subject matter of the social sciences is the reality of human experience but that scientific studies of social behavior and institutions tended to sacrifice the larger facts of human experience to the logic of analytical categories. He believed that social scientists should strive for objectivity and analytical precision, but he argued that they should also return again and again to the existential reality. In this way their research would remain relevant to their subject matter.

CHARLES M. LESLIE

[Directly related are the entries CULTURE, article on CULTURE CHANGE; ECONOMIC ANTHROPOLOGY; MIDDLE AMERICAN SOCIETY; PEASANTRY; TRIBAL SO-

CITY; VALUES, article on VALUE SYSTEMS; VILLAGE; WORLD VIEW. Other relevant material may be found in the biographies of BECKER; DURKHEIM; MAINE; MALINOWSKI; PARK; RADCLIFFE-BROWN; TÖNNIES; WIRTH.]

#### WORKS BY REDFIELD

- 1930 *Tepoztlán, a Mexican Village: A Study of Folk Life*. Univ. of Chicago Press.  
 (1934) 1962 REDFIELD, ROBERT; and VILLA ROJAS, ALFONSO *Chan Kom: A Maya Village*. Carnegie Institution of Washington, Publication No. 448. Univ. of Chicago Press.  
 1941 *The Folk Culture of Yucatan*. Univ. of Chicago Press.  
 1950 *A Village That Chose Progress: Chan Kom Revisited*. Univ. of Chicago Press. → A paperback edition was published in 1962.  
 1953 *The Primitive World and Its Transformations*. Ithaca, N.Y.: Cornell Univ. Press. → A paperback edition was published in 1957.  
 1955 *The Little Community: Viewpoints for the Study of a Human Whole*. Univ. of Chicago Press. → Paperback editions, bound together with *Peasant Society and Culture*, were published in 1960 by the Univ. of Chicago Press and in 1961 by the Cambridge Univ. Press.  
 1956 *Peasant Society and Culture: An Anthropological Approach to Civilization*. Univ. of Chicago Press.  
 1962-1963 *Papers*. 2 vols. Edited by Margaret P. Redfield. Univ. of Chicago Press. → Volume 1: *Human Nature and the Study of Society*. Volume 2: *The Social Uses of Social Science*.

#### SUPPLEMENTARY BIBLIOGRAPHY

- COLE, FAY-COOPER; and EGGAN, FRED 1959 Robert Redfield: 1897-1958. *American Anthropologist* New Series 61:652-662. → Includes eight pages of bibliography.

#### REFERENCE BOOKS

See under INFORMATION STORAGE AND RETRIEVAL.

#### REFERENCE GROUPS

Sociologists, social psychologists, and cultural anthropologists have always operated on the fundamental principles that an individual's attitudes and conduct are shaped by the group in which he has membership and that self-appraisal and the correlative feelings and behavior flow from the individual's location in a particular group within a social hierarchy. The evidence in support of such principles is, indeed, abundant; but, at times, faith in the principles becomes shaky in the face of contradictory examples: upper-class individuals with radical ideologies and revolutionary allegiances, those who feel deprived despite relatively advantaged positions, the products of an orthodox milieu who end up as nonconformists. Ordinary language is rich in terms that describe such individuals who



do not show the stamp of their group: the renegade Catholic, the genteel poor, the Tory worker.

Through the concept *reference group*, our confidence in the fundamental principles has been restored and theory and research on group influences has been invigorated. In the process of self-appraisal, from many possible groups available as a framework for social comparison, individuals make their own particular selections, thus reflecting the true complexities of their social location but not necessarily the arbitrary social position to which the scientist may have assigned them. In shaping their attitudes men may orient themselves to groups other than their own. If the groups to which individuals refer themselves, their reference groups, are empirically determined, knowledge and predictions of attitude, self-evaluation, and conduct will be enhanced; the cherished principles about group influences can be protected; and an understanding of the complex processes by which men relate themselves to groups can be enriched. Such is the hope of reference group theory and research.

The fact that men may shape their attitudes by reference to groups other than their own and may shape their self-evaluations by the choice of unusual points of social comparison is perhaps the most distinctive contribution of reference group theory. To be sure, anomalous patterns of behavior may be understood without recourse to the concept. Some members of a group may depart from the modal pattern of behavior simply because of their simultaneous membership in other groups. Some individuals in a particular status may have an incongruent self image because they occupy other statuses as well, and the "status-set," rather than the discrete status, governs the process. But even here the concept of reference group makes a distinctive contribution to what otherwise would remain problematical: Which of the multiple memberships and multiple statuses weighs heaviest with the individual, and what weights best represent their respective contributions?

### History of the concept

The term "reference group" was first used by Hyman, who elaborated the concept and explored some of its properties in "The Psychology of Status" (1942). Seeking to understand the ways in which individuals ranked themselves in terms of their choice of a social framework for comparison, he first explored, by interview, the reference groups and reference individuals that subjects employed and some of the dynamics underlying such selection, and then used experimental manipulation to determine the effects of particular reference groups

on self-appraisal. At about the same time, Newcomb (1943), searching to understand processes of attitude change, or lack of change, among individuals, systematically explored (by interview and repeated testing) the various ways in which students at Bennington College related themselves to the college community—in other words, chose this community as a reference group. These first systematic studies by social psychologists stimulated a few others, notably the Hartleys and Sherif, to continue research on reference groups. Sherif emphasized reference groups in his *Outline of Social Psychology* (see Sherif 1948), which included a summary by Newcomb of his Bennington study, rephrased in terms of the explicit concept *reference group*.

The concept had clarified for Newcomb various paradoxical findings. Similarly, Stouffer and his associates were led in their study *The American Soldier* (1949, vol. 1) to the concept of *relative deprivation*, a close cousin to the concept of comparative reference group, as they confronted the apparent contradictions between feelings of satisfaction or deprivation and the objective situation among groups of soldiers. They then invoked the interpretive principle that the soldier's sense of deprivation was not dependent on any absolute level but was relative to the perceived level in the groups with which he compared himself. These ideas and concepts, however, had little prominence until Merton and Kitt (1950) synthesized and presented them in systematic form.

Certainly, in the past, one can find precursors of the ideas implicit in reference group theory. William Graham Sumner's idea of in-groups and out-groups is a distant relation; C. H. Cooley's discussion of selective affinity to groups outside of one's immediate environment is an even earlier and closer relative (see Merton & Kitt 1950, pp. 101–102; Merton 1957, pp. 358–359). Cooley's notion of imaginary conversation with an "interlocutor" anticipates the concept of *reference individual* and has inspired recent research which applies reference group concepts to mass communication (Bauer 1958). Cooley's remark that "people differ much in the vividness of their imaginative sociability" (1902, p. 95 in the 1964 edition) is suggestive of later findings on individual differences in the use of multiple reference groups. In 1890, William James, in his account of the "social self," suggested that our *potential* social self is developed and inwardly strengthened by thoughts of remote groups and individuals who function as normative points of reference.

Since 1950 the concept has figured in so many

writings that the more recent history defies brief review. The concept appears in Australia and Israel; in studies of farmers, scientists, drunkards, and newspapermen; it has been applied to problems of mental illness, formal organization, marketing and public relations, mass communication, acculturation, political behavior, consumer behavior, and juvenile delinquency, as well as to opinion formation. This sketchy sampling conveys the wild growth, but it should also be noted that despite the general flowering, some branches have not flourished. Taking as a comprehensive outline Merton's formulation that "... reference group theory aims to systematize the determinants and consequences of those processes of evaluation and self-appraisal in which the individual takes the values or standards of other individuals and groups as a ... frame of reference" (Merton & Kitt 1950, pp. 50-51), the deficiencies as well as the accomplishments will become apparent.

### Clarification of concepts

Kelley's distinction (1952) between *comparative* and *normative* reference groups is basic; it corresponds to the two functions of reference groups as standards of comparison for self-appraisal and as the source of the individual's norms, attitudes, and values. These two functions of reference groups may be conceived of as separate but equal in importance for study, having only the common property that the individual's choice of a point of reference is the key to understanding both the process of self-appraisal and the formation of attitudes. The two types, however, may not always be empirically distinct; indeed, one group can serve both functions. Contained within the structure of norms in a group may be the directive that one should not compare himself with his betters, or look down on his inferiors, or even be aware of their existence. Indeed, in Hyman's interviews (1942) some subjects claimed that they did not employ any comparative reference group whatsoever, because of ideological distaste, and the comparative groups that other subjects employed were clearly shaped by their political attitudes. Given the possible interdependence of the two types of processes, it is all the more strange that while the study of the normative reference group has been cultivated, that of the comparative reference group has been neglected. The paths that Hyman, the Hartleys, Stouffer, and Merton took are now only byways, trodden by occasional investigators (see Patchen 1961; Runciman 1961; Form & Geschwender 1962).

The equally basic distinction between reference

*individuals* and *reference groups* has been neglected despite the emphasis on reference individuals as points of social comparison in the early work and the distinction's connection to such prestigious concepts as "role-model" and the current study of personal influence and opinion leadership (see Merton 1957, pp. 302-304; Hyman 1960). The parenthetical remark by Newcomb (1952, p. 420) that "individuals acquire the approved attitudes to the extent that the membership group (particularly as symbolized by leaders . . .) serves as a positive point of reference" strongly suggests the role of the reference individual as the carrier of the reference group's norms, but it appears to have been neglected by subsequent researchers.

For the study of normative reference groups, Newcomb's distinction between the *positive* and *negative* types reminds us that individuals may form their attitudes in opposition to the norms of a group toward which they are negative. The concept of negative reference groups helps us to understand not only the affective tone and content of an individual's attitude but also such *formal* features as the congruence and organization of his attitudes. Clearly, there are some instances in social life where to oppose the norms of a particular group—for example, the Republican party—is to be thrown into the arms of its opposite, the Democratic party. But there are many other instances where social relations between groups are not patterned in terms of polar opposites. Thus, to regard one's parents or community as a negative reference group may provide no other directive to the individual than to choose from among the norms of the myriad groups available. Individuals who form a constellation of attitudes under such conditions may well show the consequences in terms of diffuseness of attitudes, lack of crystallization, inconsistency, etc.; however, this hypothesis remains to be tested.

The concept *referent power*, employed by French and Raven (1959), suggests many fundamentals of normative reference group processes. The power of a nonmembership reference group inheres essentially in the fact that the individual, by his sheer identification with the group, *willingly* accepts what he perceives to be its norms. By contrast, membership groups often have the power, even when the individual does not take them as reference groups, to exact conformity in behavior through brutish means or rewards and to induce attitudes through prolonged doses of socialization. Certainly, when there is no bond of identification, their influence may be attenuated, and the concept of reference group reminds us that the psychological equipment of an individual can provide some escape from



victimization by a membership group. Of course, when referent power is joined to real power, the combination is unbeatable. From these distinctions flow hypotheses that the attitude *held* reflects the reference group, whereas the attitude *expressed* reflects the membership group (see Smith et al. 1956).

Merton's concept of *anticipatory socialization* is essential to this discussion (see Merton & Kitt 1950, pp. 87-88). Individuals may take as a reference group a nonmembership group to which they aspire to belong, and begin to socialize themselves to what they perceive to be its norms before they are ever exposed to its influence. The power of some reference groups thus inheres in the fact that they will ultimately be membership groups—at least such is the belief of the aspirant—and therefore can exact some conformity as the price of admission or of more comfortable passage into their ranks. Eulau (1962) advanced and then twice tested an ingenious hypothesis: that anticipatory socialization may be an effective means for learning *attitudes*, but not *conduct*, since the aspirant will have had little real opportunity to practice the skills required and to be taught the correct performance of the role.

Basic to reference group theory is the fact that individuals often have *multiple* reference groups. Certainly, there are some individuals who have limited capacity to use many reference groups—who lack rich “imaginative sociability.” Others, however, in appraising the many facets of the self, employ various reference groups, each specialized as a point of comparison for one particular dimension (Hyman 1942; Stouffer et al. 1949; Turner 1955). In forming the total constellation of attitudes, several reference groups may be employed, each accorded a limited jurisdiction over some specialized attitude sphere. Studies of normative reference groups have found differences in the legitimacy that individuals accord to groups promulgating norms in various spheres (Michigan . . . 1960). There are also instances where multiple reference groups impinge simultaneously on the same sphere of comparison or the same realm of attitude, and then they may either reinforce the same outcome or produce conflicting consequences for the individual (see Stouffer et al. 1949; Merton & Kitt 1950; Form & Geschwender 1962; Patchen 1961; Rosen 1955).

Over the life span of any person, there will have been a multiplicity of reference groups, specialized less by sphere than by the life segment to which they were keyed. Some old reference groups which are long departed may be carried over in memory.

On the other hand, recent groups may be cast out of mind in the zealous adoption of a still newer reference group. The relations of multiple reference groups within a sequence suggest many fascinating problems that tie into the processes of social mobility (see Merton & Kitt 1950, pp. 84-95). Discussions of social mobility often assume that the past and future reference groups conflict, since the individual presumably wishes to break his ties to the old, inferior group. Litwak, however, presented an interesting reformulation, using the concept of the “stepping-stone” reference orientation (1960). In a situation characterized by ordered change, “where integration into one group is considered to be a prerequisite for integration into a second group . . . it is possible for the individual to view both his current membership group and his future membership group as reference groups, without endangering his integration into his current group and without preventing his joining a different future group” (*ibid.*, pp. 72-73). Each group is valued by the individual as a steppingstone to help him in his advance.

The concepts reviewed in no way exhaust the literature but are those basic to any clarification of the field. Merton's essay (1957) describes many other conceptual refinements and the extended network of connections to other branches of theory (see also Shibutani 1955; Turner 1956; Hyman 1960).

### Selection of reference groups

The concept of reference group has always implied that one cannot make arbitrary assumptions about the groups to which an individual refers himself. Given the multiplicity of groups and the variability among individuals and situations, must we then, as a symposium (Foundation . . . 1956, p. 1) put it, “determine which kinds of groups are likely to be referred to by which kinds of individuals under which kinds of circumstances in the process of making which decisions . . .”—over and over again? There will always be a large amount of empiricism needed, and the development of simple instruments to measure a person's reference groups is of great importance. But, fortunately, research has already established certain regularities in the choices individuals make and some major factors governing selection.

Theorizing about the choice of reference groups and reference individuals is often based on simple assumptions about motivation. For example, the individual chooses a normative reference group so that in fantasy, or ultimately in fact, he can feel himself part of a more favored group. Or, facing

rapid social change, the individual latches onto a reference group; thus anchored, he has a ready-made perspective to order the distressing complexities of the environment (see Shibutani 1955; Eisenstadt 1954). For social comparisons, he chooses a group so as to enhance his self-regard or protect his ego. Certainly in the search for reference groups such fundamental strivings play an important part.

The pleasure principle is at work, but so, too, is the reality principle: consider, for example, Stouffer's inference that the *more* advantaged soldiers felt deprived because they chose to compare themselves with others who were even better off (Stouffer et al. 1949). It seems plausible that the institutional arrangements gave such sharp definition and prominence to certain groups that the soldiers' attention was drawn to them as points of comparison. Perhaps when reality is less highly structured, there is more freedom for the pleasure principle to guide the selection of reference groups.

The principle of similarity. Turner hypothesized that only those groups will be taken as points of comparison which are *relevant* to a particular aspect of self-appraisal; when a group's standing is so high or so low that it is not meaningful to the individual, it will not be used as a comparative reference group. The similarity principle derived by Festinger (1954) in his theory of social comparison processes—that an individual chooses others who are close to his level of ability—is congruent with Turner's "relevance principle," as is Merton's hypothesis (based on findings in *The American Soldier*) that some similarity in status attributes between the individual and the reference group must be perceived or imagined in order for the comparison to occur at all (Merton & Kitt 1950, p. 61).

The Amba of east Africa dramatically illustrate this principle. They worked for Europeans for a much lower price than for employers from another tribe, and "are quite willing to explain this state of affairs. They say that a European is on a much higher social plane, and therefore, comparisons are out of the question. Europeans are so wealthy that an increase in their wealth makes no difference in the . . . standing" of the Amba relative to Europeans (Winter 1955, p. 40). Qualitative evidence in Hyman's interviews also suggested the operation of the similarity principle, or what he called "affinity," in the choice of reference groups, but he also observed instances where *contrast* in status made a reference group salient.

The principle of relevance or similarity still leaves much room for the play of psychological

factors. *Perceived* similarity is what counts, and there are many dimensions of similarity, only some of which are noted by the individual. And inside the range of similarity, in which direction will the individual then turn—toward groups which are superior or inferior to him? Turner's college students seemed to compare themselves with higher reference groups, perhaps to their present discomfort, because they were "future-oriented," desiring to surpass such groups in their future lives (Turner 1955).

Patchen's study of industrial workers provides systematic evidence on the variables affecting the choice of a reference individual or reference group for economic comparisons, and fundamentally clarifies the motivational assumptions of reference group theory (1961, especially chapter 2). Men often choose reference groups which increase their *present* sense of relative deprivation, not only because formal institutional arrangements force such groups into attention but also, as Patchen demonstrates, because informal social influences make such groups salient. Men may choose groups above them at the price of present dissatisfaction because they are laying a claim to a future when their status will be higher and their relative deprivation diminished. Furthermore, upward comparisons are not always experienced as painful to the ego, and herein lies Patchen's singular contribution. Comparisons along a particular dimension are not made, in his words, in a "cognitive vacuum." Low rank—for example, on wages—may appear justifiable in light of the low-ranking individual's and the reference group's relatively high standing on other attributes—e.g., education or age—thus reducing the grounds for dissatisfaction. However, reference groups which are equal in respect to the relevant attributes but which are higher on the dimension of appraisal create greater feelings of deprivation. Thus the principle of similarity is specified: If the direction of choice is upward, deprivation is likely to be contained by choosing a group whose dissimilarity in other attributes legitimizes the present inferior position of the person making the comparison. However, if one selects reference groups *similar* on these attributes, the direction of comparison is not likely to be upward.

**Normative reference groups.** Research on the selection of *normative* reference groups is meager. The proposition that individuals identify with advantaged groups and thereby gain gratification must be qualified in terms of the societal context. Such modes of selection may be characteristic of societies with high rates of upward mobility (Merton & Kitt 1950, p. 91), or where upward mobility



is a strong value or is perceived, correctly or not, to be frequent. Comparative research is clearly required.

Experimental research demonstrates that situational factors may heighten the saliency of a membership group and increase the likelihood of its being used as a reference group whose perceived norms then affect some specific sphere. Charters and Newcomb (1952) present such evidence, but they note that the salient group sometimes functions as a *negative* reference group. Subsequent research puts these findings in doubt, and whether such situational influences have enduring effects on the choice of normative reference groups remains unknown (Kelley 1955).

That normative reference groups are chosen in the spirit of identification perhaps also needs qualification. It may be true for the individual seeking a source of norms and values and attitudes. But what about the individual seeking a system of beliefs and knowledge? He may then choose his reference group in terms of its authority or expertness and with the full awareness that he has no bond of identification. Systematic research on such determinants of choice does not exist, since the effects of reference groups on cognitive processes has been neglected, to the detriment of the sociology of knowledge. Prospects for future research are illustrated by Carlson (1952), who demonstrated differences in the effectiveness with which rural Southern Negroes dealt with syphilitic infections, depending on their reference groups, and by the report of a seminar conducted by the Foundation for Research on Human Behavior (1959), which shows that farmers who adopted better practices chose particular reference groups (see also Deutsch & Gerard 1955; Newcomb 1946).

Ruth Hartley's work represents a unique program of systematic research on psychological factors that influence the selection of a membership group as a normative reference group (1956; 1957; 1960a; 1960b; 1960c). Using a large college community, she measured the degree to which students adopted their new community as a reference group, and correlated such individual differences with other characteristics. Taking on a new reference group depends on possessing an acceptant personality pattern. A particular reference group is then likely to be chosen if it is seen as fulfilling personal needs and if there is congruity between the individual's personal values and norms and the norms and values he perceives as characteristic of the group. Thus some of the apparent effect of reference groups on the values of individuals may be

spurious, since their values were prior in time and determined the choice of the reference group.

### Perception of group norms and standing

For an individual to guide himself by a reference group requires some perception or cognition of its norms. Since reference groups may often be distant, nonmembership groups, perception of the true norms may be hazy and incorrect and not subject to any correction from the group. But even when membership groups function as reference groups, the visibility of group norms is not always high (Merton 1957, pp. 336-353) and varies, depending on one's position in the group (Chowdhry & Newcomb 1952). Comparative reference group processes also require some perception or knowledge of the standing of others on the dimensions selected for comparison.

Deviation from the objective position of a group thus may be inspired by conformity to a false norm which the individual has taken to be the true norm of the reference group, and conformity to the objective norms may be the perverse fate of a "deviant" who thinks he is flouting the norms when actually he is misperceiving them. Attempts to understand the *motivation* of conformists and deviants must make a clear distinction between these varieties and therefore must measure the norms imputed to groups.

The environmental conditions and psychological processes that aid or obstruct perception vary greatly with the nature of the reference group. Organized groups announce their views to members and to outsiders; diffusion is aided by the mass media and spread by word of mouth. However, self-appointed communicators and the diffusion process may also distort the norms that finally reach the individual. More fundamental problems of perception must also be considered. Even organized groups do not announce all of their views all of the time: a norm may not yet have been promulgated or a fundamental value may remain implicit and taken for granted. On certain issues, the norms may be confused, not shared throughout the organization, exceedingly complex in nature, or not distinctive from those of other groups. There are even occasional groups whose fundamental value is that the individual shall be autonomous in regulating his conduct, thereby creating difficulty for those individuals seeking cues. Such are the burdens on the perceiver, but what comes to his aid is time; he has abiding loyalties to some reference groups, and what at first is dimly perceived finally takes on clarity.

Not all reference groups are organized entities. They may be vague collectivities, sprawling social categories, groups out of the dead past, or groups not yet born. Such groups are living structures only in the mind of the perceiver and do not communicate or transact behavior. Here there is relatively free rein for autistic perception of norms.

The choice of reference groups from the immediate environment, or from membership groups, or the choice of *reference individuals* rather than groups, may be motivated by the individual's need to simplify his perceptual tasks. Surveys of political behavior provide relevant evidence of the choice of membership groups. Members of a given social category have greater awareness than do nonmembers that a voting norm characterizes the group; among nonmembers, awareness of the norm of such a group is greater for those whose environment contains many representatives of that group (Campbell et al. 1954; Hyman 1960). Donald Campbell reformulated the concept of the "bandwagon effect" as the influence of the voting norm of an assumed reference group, and he designed an experiment to test whether the local, state, or national reference group norm was effective (1951). But, alas, it was never carried out.

### Measurement procedures

The popularity of reference group theory brings the danger, in Sherif's words, that the concept is "becoming a magic term to explain anything and everything concerning group relations" (1953, p. 204). The concept is often invoked and the influence of a particular reference group alleged without benefit of evidence. Magical invocation must give way to scientific measurement that will lead cumulatively to knowledge on the problematics of reference group theory.

Only a brief summary of research procedures is possible in this article. A person's reference groups have been measured by such simple and yet predictive questions as those on subjective class identification (Centers 1949). When this "self-location question" is combined with a question on class awareness, prediction of attitudes is improved (Michigan . . . 1960, chapter 14). Other simple questions on the importance of a series of groups for the individual can predict his attitudes (see Hyman 1960). Strength of attachment to one of a series of possible normative reference groups has been measured by a forced-choice instrument developed by Melikian and Diab (1959). Ruth Hartley's elaborate instrument to measure the acceptance of a normative reference group has high

reliability and has been validated (1956). The type of comparative reference groups that individuals normally employ has been studied by direct questions (Patchen 1961), by the spontaneous group references that individuals make in the course of surveys on personal satisfaction (Stern & Keller 1953), and by applications of Kuhn's "Who Am I?" test, which elicits spontaneous definitions of the self and its incorporation into various social categories (see Mulford & Salisbury 1964).

Perception of the norms of a reference group is measured by having the subject estimate the opinion of various other groups or individuals or indicate his lack of awareness of any norm (Newcomb 1943; Kaplan 1955). Centers (1953) reversed this procedure by presenting subjects with statements of beliefs or norms and asking them to impute the belief to some group. Perception of the *differentiated* norms within a complex group has been examined in surveys of political behavior and in Ruth E. Hartley's battery on "perceived cohesiveness" of a reference group; the perceived clarity and perceived uniqueness of norms, as well as the legitimacy an individual accords to a reference group exercising a norm in a particular sphere, have been measured by simple batteries of questions (for a summary, see Hyman 1960). Experimental techniques to study the influence of reference groups on the behavior of a communicator have been developed by Zimmerman and Bauer (1956) and by Pool and Shulman (1959). Experiments to test the effects of manipulating such variables as the saliency of reference groups have also been designed (Charters & Newcomb 1952).

Current investigators have no excuse for neglecting existing measurement procedures. But they also have an obligation to refine these methods. As Newcomb has remarked: ". . . the concept of reference group is almost unique among the tools available to the social psychologist . . . It is a variable intimately associated with that central problem of social psychology: the relating of self to society. The hand-to-hand advancement of reference-group theory and of the research procedures which can make it possible would therefore seem to be one of social psychology's greatest needs" (1951, p. 92).

HERBERT H. HYMAN

[Directly related are the entries ATTITUDES; IDENTITY, PSYCHOSOCIAL; SOCIAL PSYCHOLOGY; SOCIALIZATION; STATUS, SOCIAL. Other relevant material may be found in CASTE, article on THE CONCEPT OF CASTE; DELINQUENCY, article on THE STUDY OF DELIN-



QUENCY; and in the biographies of COOLEY; JAMES; STOUTER; SUMNER.]

## BIBLIOGRAPHY

- BAUER, RAYMOND A. 1958 The Communicator and the Audience. *Journal of Conflict Resolution* 2:67-77.
- BENNIS, W. G. et al. 1958 Reference Groups and Loyalties in the Out-patient Department. *Administrative Science Quarterly* 2:481-500.
- CAMPBELL, ANGUS; GURIN, GERALD; and MILLER, WARREN E. 1954 *The Voter Decides*. Evanston, Ill.: Row, Peterson.
- CAMPBELL, DONALD T. 1951 On the Possibility of Experimenting With the "Bandwagon" Effect. *International Journal of Opinion and Attitude Research* 5: 251-260.
- CARLSON, ROBERT O. 1952 The Influence of the Community and the Primary Group on the Reactions of Southern Negroes to Syphilis. Ph.D. dissertation, Columbia Univ.
- CENTERS, RICHARD (1949) 1961 *The Psychology of Social Classes: A Study of Class Consciousness*. New York: Russell.
- CENTERS, RICHARD 1953 Social Class, Occupation, and Imputed Belief. *American Journal of Sociology* 58: 543-555.
- CHARTERS, W. W. JR.; and NEWCOMB, THEODORE M. 1952 Some Attitudinal Effects of Experimentally Increased Salience of a Membership Group. Pages 415-420 in *Society for the Psychological Study of Social Issues, Readings in Social Psychology*. Rev. ed. Edited by Guy E. Swanson, Theodore M. Newcomb, and Eugene L. Hartley. New York: Holt.
- CHOWDHRY, KAMLA; and NEWCOMB, THEODORE M. 1952 The Relative Abilities of Leaders and Non-leaders to Estimate Opinions of Their Own Groups. *Journal of Abnormal and Social Psychology* 47:51-57.
- COOLEY, CHARLES H. (1902) 1956 *Human Nature and the Social Order*. Rev. ed. In Charles H. Cooley, *Two Major Works: Social Organization and Human Nature and the Social Order*. Glencoe, Ill.: Free Press. → Each title reprinted with individual title page and pagination. Separate paperback editions were published in 1964 by Schocken.
- DEUTSCH, MORTON; and GERARD, HAROLD B. 1955 A Study of Normative and Informational Social Influences Upon Individual Judgment. *Journal of Abnormal and Social Psychology* 51:629-636.
- DUESENBERY, JAMES S. 1949 *Income, Saving and the Theory of Consumer Behavior*. Harvard Economic Studies, Vol. 87. Cambridge, Mass.: Harvard Univ. Press.
- EISENSTADT, SHMUEL N. 1954 Studies in Reference Group Behavior. I. Reference Norms and the Social Structure. *Human Relations* 7:191-216.
- EULAU, HEINZ 1962 *Class and Party in the Eisenhower Years: Class Roles and Perspectives in the 1952 and 1956 Elections*. New York: Free Press.
- FESTINGER, LEON 1954 A Theory of Social Comparison Processes. *Human Relations* 7:117-140.
- FORM, WILLIAM H.; and GESCHWENDER, JAMES A. 1962 Social Reference Basis of Job Satisfaction: The Case of Manual Workers. *American Sociological Review* 27:228-237.
- FOUNDATION FOR RESEARCH ON HUMAN BEHAVIOR 1956 *Group Influence in Marketing and Public Relations: Report of a Seminar*. Ann Arbor, Mich.: The Foundation.
- FOUNDATION FOR RESEARCH ON HUMAN BEHAVIOR 1959 *The Adoption of New Products: Process and Influence*. Ann Arbor, Mich.: The Foundation.
- FRENCH, JOHN R. P.; and RAVEN, BERTRAM 1959 The Bases of Social Power. Pages 150-167 in Dorwin Cartwright (editor), *Studies in Social Power*. Research Center for Group Dynamics, Publication No. 6. Ann Arbor: Univ. of Michigan, Institute for Social Research.
- HARTLEY, EUGENE L. 1951 Psychological Problems of Multiple Group Membership. Pages 371-387 in Conference on Social Psychology at Crossroads, University of Oklahoma, 1950, *Social Psychology at the Crossroads*. Edited by John H. Rohrer and Muzafer Sherif. New York: Harper.
- HARTLEY, RUTH E. 1956 The Acceptance of New Reference Groups. Technical Report No. 3. Unpublished manuscript, U.S. Office of Naval Research.
- HARTLEY, RUTH E. 1957 Personal Characteristics and Acceptance of Secondary Groups as Reference Groups. *Journal of Individual Psychology* 13:45-55.
- HARTLEY, RUTH E. 1960a Norm Compatibility, Norm Preference, and the Acceptance of New Reference Groups. *Journal of Social Psychology* 52:87-95.
- HARTLEY, RUTH E. 1960b Personal Needs and the Acceptance of a New Group as a Reference Group. *Journal of Social Psychology* 51:349-358.
- HARTLEY, RUTH E. 1960c Relationships Between Perceived Values and Acceptance of a New Reference Group. *Journal of Social Psychology* 51:181-190.
- HASKELL, MARTIN R. 1960-1961 Toward a Reference Group Theory of Juvenile Delinquency. *Social Problems* 8:220-230.
- HYMAN, HERBERT H. 1942 The Psychology of Status. *Archives of Psychology* No. 269.
- HYMAN, HERBERT H. 1960 Reflections on Reference Groups. *Public Opinion Quarterly* 24:383-396.
- HYMAN, HERBERT H.; and SINGER, ELEANOR (editors) 1968 *Readings in Reference Group Theory and Research*. New York: Free Press.
- JAMES, WILLIAM (1890) 1962 *The Principles of Psychology*. 2 vols. New York: Smith. → See especially Volume 1.
- KAPLAN, NORMAN 1955 Reference Group Theory and Voting Behavior. Ph.D. dissertation, Columbia Univ.
- KELLEY, HAROLD H. 1952 Two Functions of Reference Groups. Pages 410-414 in *Society for the Psychological Study of Social Issues, Readings in Social Psychology*. Rev. ed. Edited by Guy E. Swanson, Theodore M. Newcomb, and Eugene L. Hartley. New York: Holt.
- KELLEY, HAROLD H. 1955 Salience of Membership and Resistance to Change of Group-anchored Attitudes. *Human Relations* 8:275-289.
- LITWAK, EUGENE 1960 Reference Group Theory, Bureaucratic Career, and Neighborhood Primary Group Cohesion. *Sociometry* 23:72-84.
- MELIKIAN, LEVON H.; and DIAB, LUTFY N. 1959 Group Affiliations of University Students in the Arab Middle East. *Journal of Social Psychology* 49:145-159.
- MERTON, ROBERT K. 1957 Continuities in the Theory of Reference Groups and Social Structure. Pages 281-386 in Robert K. Merton, *Social Theory and Social Structure*. 2d ed., rev. & enl. Glencoe, Ill.: Free Press.
- MERTON, ROBERT K.; and KITT, ALICE S. 1950 Contributions to the Theory of Reference Group Behavior. Pages 40-105 in Robert K. Merton and Paul F. Lazarsfeld (editors), *Continuities in Social Research: Studies*

- in the *Scope and Method of The American Soldier*. Glencoe, Ill.: Free Press.
- MICHIGAN, UNIVERSITY OF, SURVEY RESEARCH CENTER 1960 *The American Voter*, by Angus Campbell et al. New York: Wiley.
- MULFORD, HAROLD A.; and SALISBURY, WINFIELD W. 1964 Self-conceptions in a General Population. *Sociological Quarterly* 5:35-46.
- NEWCOMB, THEODORE M. (1943) 1957 *Personality and Social Change: Attitude Formation in a Student Community*. New York: Dryden.
- NEWCOMB, THEODORE M. 1946 The Influence of Attitude Climate Upon Some Determinants of Information. *Journal of Abnormal and Social Psychology* 41: 291-302.
- NEWCOMB, THEODORE M. 1951 [A Book Review of] *Continuities in Social Research: Studies in the Scope and Method of The American Soldier*, edited by Robert K. Merton and Paul F. Lazarsfeld. *American Journal of Sociology* 57:90-92.
- NEWCOMB, THEODORE M. 1952 Attitude Development as a Function of Reference Groups: The Bennington Study. Pages 420-430 in *Society for the Psychological Study of Social Issues, Readings in Social Psychology*. 2d ed. Edited by Guy E. Swanson, Theodore M. Newcomb, and Eugene L. Hartley. New York: Holt.
- PATCHEN, MARTIN 1961 *The Choice of Wage Comparisons*. Englewood Cliffs, N.J.: Prentice-Hall.
- POOL, ITHIEL DE SOLA; and SHULMAN, IRWIN 1959 *Newsmen's Fantasies, Audiences, and Newswriting*. *Public Opinion Quarterly* 23:145-158.
- REICHENBACH, HANS 1951 *Probability Methods in Social Science*. Pages 121-128 in Daniel Lerner and Harold D. Lasswell (editors), *The Policy Sciences: Recent Developments in Scope and Method*. Stanford Univ. Press.
- ROGERS, EVERETT M.; and BEAL, GEORGE M. 1958 *Reference Group Influences in the Adoption of Agricultural Technology*. Ames: Iowa State College, Department of Economics and Sociology.
- ROSEN, BERNARD C. 1955 *Conflicting Group Membership: A Study of Parent-Peer Group Cross-pressures*. *American Sociological Review* 20:155-161.
- RUNCIMAN, W. G. 1961 Problems of Research on Relative Deprivation. *Archives européennes de sociologie* 2:315-323.
- SHERIF, MUZAFA 1953 The Concept of Reference Groups in Human Relations. Pages 203-231 in *Conference in Social Psychology*, University of Oklahoma, 1952, *Group Relations at the Crossroads*. Edited by Muzafer Sherif and M. O. Wilson. New York: Harper.
- SHERIF, MUZAFA; and SHERIF, CAROLYN W. (1948) 1956 *An Outline of Social Psychology*. Rev. ed. New York: Harper. → Muzafer Sherif was the sole author of the first edition.
- SHIBUTANI, TAMOTSU 1955 Reference Groups as Perspectives. *American Journal of Sociology* 60:562-569.
- SMITH, M. BREWSTER; BRUNER, JEROME S.; and WHITE, ROBERT W. 1956 *Opinions and Personality*. New York: Wiley.
- STERN, ERIC; and KELLER, SUZANNE 1953 Spontaneous Group References in France. *Public Opinion Quarterly* 17:208-217.
- STOUFFER, SAMUEL A. et al. 1949 *The American Soldier*. Studies in Social Psychology in World War II. Vols. 1 and 2. Princeton Univ. Press. → Volume 1: *Adjustment During Army Life*. Volume 2: *Combat and Its Aftermath*. See especially Volume 1.
- TAFT, RONALD 1952 Minority Group Behavior and Reference Group Theory: A Reply to Emery and Katz. *Australian Journal of Psychology* 4:10-23.
- TURNER, RALPH H. 1955 Reference Groups of Future Oriented Men. *Social Forces* 34:130-136.
- TURNER, RALPH H. 1956 Role-taking, Role Standpoint, and Reference-group Behavior. *American Journal of Sociology* 61:316-328.
- WINTER, E. H. 1955 *Bwamba Economy: The Development of a Primitive Subsistence Economy in Uganda*. East African Institute of Social Research, East African Studies, No. 5. Kampala (Uganda): The Institute.
- ZIMMERMAN, CLAIRE; and BAUER, RAYMOND A. 1956 The Effect of an Audience Upon What Is Remembered. *Public Opinion Quarterly* 20:238-248.

## REFERENDUM

See ELECTIONS; REPRESENTATION.

## REFUGEES

- |                                 |                   |
|---------------------------------|-------------------|
| I. WORLD PROBLEMS               | Louise W. Holborn |
| II. ADJUSTMENT AND ASSIMILATION | Judith T. Shuval  |

### I

#### WORLD PROBLEMS

The refugee problem is a phenomenon of our age. It is the product not only of the most destructive wars of history, World War I and World War II, of modern dictatorial regimes, and of the national awakening of peoples, but also of the closed frontiers characteristic of the twentieth century. There were refugees in earlier centuries but no refugee problem in the modern sense, for the involuntary migrant could merge with those who by choice sought new homes elsewhere. In our time, the refugee problem has been distinguished from refugee movements of earlier days by its scope, variety of causes, and difficulty of solution.

Modern refugee movements, beginning in Europe and subsequently becoming world-wide, have given rise to a new class of people who are homeless and stateless and who live in a condition of constant insecurity which erodes human dignity. They have caused grave political and economic problems for the countries of temporary reception, problems which have proved too burdensome for the administrative facilities and financial resources of private organizations and national governments. The refugee problem has thus transcended national jurisdiction and institutions.

Furthermore, while in its earlier stages the refugee problem was seen as a temporary and limited phenomenon, it has now come to be acknowledged as universal, continuing, and recurring. In response to this challenge the international community has developed a complex mechanism



of world-wide cooperation involving a tripartite partnership of national governments, private agencies, and international organizations; no longer confined by strict definitions of the word "refugee," it is prepared to approach the problem in all its various aspects—political, social, economic, and humanitarian.

### Defining the refugee

There is no single definition of "refugee" that is suitable for all purposes. When associated with humanitarian aims, the connotation of the term differs from that used in international agreements, since the human aspects of the refugee problem are clearly distinct from the question of a refugee's status in any given situation (Rees 1957; Weis 1960). However, all refugees have in common these characteristics: they are uprooted, they are homeless, and they lack national protection and status.

The refugee is an involuntary migrant, a victim of politics, war, or natural catastrophe. Every refugee is naturally a migrant, but not every migrant is a refugee. A migrant is one who leaves his residence (usually for economic reasons) in order to settle elsewhere, either in his own or in another country. A refugee movement results when the tensions leading to migration are so acute that what at first seemed to be a voluntary movement becomes virtually compulsory. The uprooted become either internal refugees, that is, "national refugees" (persons who have been displaced in their own country), or "international refugees" (persons outside their country of origin). The latter are designated refugees in legal terminology when they lack the diplomatic protection granted to nationals abroad.

There are no generally accepted criteria to determine when a refugee ceases to be a refugee. Proposed criteria include: when the refugee is earning a living and has found a permanent place to live; when he has acquired a new nationality and has obtained equal rights with the inhabitants of the country of asylum or resettlement; and both criteria together.

Statistics on refugees are affected by many factors: the difficulty of obtaining accurate data; the purpose for which the statistics are compiled; and the fact that refugees are always on the move. The best estimate for 1967—over 11 million refugees—is based on persons registered by the Office of the United Nations High Commissioner for Refugees (UNHCR), voluntary agencies, and governmental and intergovernmental offices. Doubtless there are many unrecorded refugees receiving no

aid; it is equally likely that many instances of duplicated registration occur, enabling some refugees to benefit from more than one agency.

### European refugee movements

Europe, in the twentieth century, has been a vast sea of refugee movements set in motion by the disruptions of war, the breakup of empires, the impact of violent nationalism, and the arbitrary actions of dictatorial regimes. Early in the century, political turbulence in the Balkans and Asia Minor resulted in the movement of hundreds of thousands of people from one country to another, culminating in large exchanges of populations, in particular of Greeks, Bulgars, Serbs, Armenians, and Turks. The Convention of Lausanne, January 30, 1923, stipulated a compulsory exchange of populations between Greece and Turkey. A total of 1.3 million Greeks, including tens of thousands from Russia and Bulgaria, were transferred to Greece, and about 400,000 Turks to Turkey. Between 1913 and 1925 more than 220,000 Bulgars moved into the truncated territories of Bulgaria, and the Convention of Neuilly, November 27, 1919, provided for a voluntary exchange of populations between Greece and Bulgaria. After the Russian collapse in 1917, 30,000 Assyrians who had fought against the Turks escaped to the Caucasus, Greece, and Iraq, and later to Syria. Armenians fled from persecution and massacre in Asia Minor following the collapse of the Ottoman Empire and the rise of Turkish nationalism. By 1923 an estimated 320,000 Armenian refugees were scattered in the Middle East, the Balkans, and other European countries (Simpson 1939a, pp. 11-61).

About 1.5 million Russian nationals were dispersed and left stranded in north, central, and southern Europe and in the Far East as a result of the Bolshevik Revolution of November 1917, the rout of the anti-Bolshevik armies in European Russia in 1919-1920, the famine of 1921, and the collapse of White Russian resistance in Siberian Russia in 1922 (Kulischer 1943, p. 39; Simpson 1939a, pp. 62-125).

The ranks of refugees were further increased by those in flight from dictatorships in Spain, Germany, and Italy. Some 140,000 of the Spanish refugees who sought safety in France between 1937 and 1939 remained in that country after the Spanish Civil War came to an end; smaller numbers, especially children, were evacuated to Great Britain, Belgium, Mexico, and the Soviet Union; between 40,000 and 50,000 fled to north Africa (Simpson 1939b, pp. 58-63). Between 1933 and

the outbreak of World War II, more than a million refugees, most of them Jews, left Germany: many succeeded in fleeing to western Europe or across the seas, but nearly 700,000 of them remained in the territories subsequently occupied by Germany and its allies. Refugees from Italian dictatorship numbered 65,000–70,000 in 1938; many of them went to north Africa (Simpson 1939a, pp. 117–125).

**Refugees in World War II.** World War II caused the most formidable displacement of population ever experienced. First, there was the mass movement of Germans within "Greater Germany." Ethnic Germans were transferred into Germany, mainly from eastern Europe; it has been estimated that approximately 600,000 persons had been transferred into the German Reich by the spring of 1942 (Kulischer 1943, p. 25). Other government measures for the movement of German nationals included the dispersal of industry in Germany, the "colonization of the conquered territories," and evacuation from bomb-target urban centers. In addition, hundreds of thousands of German Jews were herded into concentration and extermination camps. Finally, large numbers of Germans fled from the path of the victorious Allied armies. Then there was the displacement of non-Germans: those expelled from the defeated countries; those whose movement was effected by agreements or treaties for the transfer and exchange of populations; those dispatched to "Greater Germany" as prisoners of war or forced laborers; and those non-Germans, mostly Jews, systematically deported from the defeated countries to the concentration camps of Germany. It has been estimated that by May 1945 there were 40.5 million uprooted people in Europe, excluding non-German forced laborers and those Germans who fled before the advancing Soviet armies (Kulischer 1948, pp. 255–273).

**Postwar refugees from eastern Europe.** About 1.6 million persons displaced from east European countries during the period 1939–1945 are estimated to have refused repatriation after World War II (these figures are based principally on those compiled by the Occupation authorities in postwar Germany and by the International Refugee Organization). But the first major postwar movement of refugees from eastern Europe was as a result of the communist coup in Czechoslovakia in February 1948, when 60,000 Czech refugees fled to the western zones of Germany and Austria. At the outbreak of the Hungarian revolution in October 1956, more than 200,000 Hungarian refugees poured over the borders into Austria (180,000) and Yugoslavia (20,000). There is

still a small but steady flow westward of escapees from the communist countries of eastern Europe, estimated, at the end of 1964, to be between 12,000 and 15,000 per year. During the period 1945–1966, according to records kept by the U.S. Escapee Program, a total of about 1,270,000 persons escaped from eastern European countries to western European countries. West German authorities estimate that, during the same period, some 3,735,000 German refugees fled into West Germany from East Berlin and other parts of East Germany. A similar number are escaping from Communist China to other Asian countries.

**"Repatriates" to European countries.** The 9.7 million German "expellees" from the territories east of the Oder and Neisse rivers, the 3.3 million persons from eastern and southern Europe, and the nearly 65,000 persons who moved, under governmental agreement for family reunion, from East Germany to West Germany, were national and ethnic German refugees for whom the West German government assumed responsibility (Holborn 1956, p. 15; International Labor Office 1959, chapter 1). Similar problems attended the liquidation of European colonial regimes. For instance, the years 1945–1958 saw the repatriation of many French citizens from former French territories in Asia and Africa: 75,000 returned from Indochina after the establishment of independent Vietnam, 10,000 left both before and after the establishment of the Republic of Guinea in 1958, and 138,000 came from Tunisia and 172,000 from Morocco after the attainment of independence by those countries in 1956. Over 15,000 French citizens left Egypt after the Suez crisis in 1956. Some 950,000 more left Algeria during and after the struggle for independence which culminated in the establishment of an independent republic in 1962 (McDonald 1965). The Netherlands received about 300,000 Dutch Indonesians when Indonesia became an independent state in 1949 (Smith 1966, p. 47).

All of the above were accepted by their respective countries as national refugees, i.e., as legal citizens. The relationship between the repatriates and the indigenous population was very strained at first, particularly in regard to employment and social adjustment, but the national governments assisted in the settlement and integration of the newcomers and, in the case of the Dutch Indonesians, in the resettlement of many overseas.

### Refugees in Asia

In the second half of the twentieth century, the scene of mass population movements has shifted



from Europe to Asia. The people involved have tended to come from a simple agricultural setting, and their flight, for the most part, has been to equally undeveloped and often politically unstable countries. The result has been widespread destitution and misery.

**India and Pakistan.** Following the partition of British India, in 1947, into the two sovereign states of India and Pakistan, 15 million people crossed the newly defined borders, in the greatest mass migration ever recorded. A government refugee program was evolved for the estimated 8 million Hindus who, by 1954, had crossed over into India; and the Pakistan government reported that a similar number of Muslim refugees from India had to be provided for by government programs (U.S. Congress . . . 1954; International Labor Office 1959, pp. 108-120). A further uprooting, though on a much smaller scale, took place as a result of the hostilities between India and Pakistan over Kashmir in August 1965.

**Korea.** In the wake of the defeated Japanese armies, U.S. and Soviet forces occupied Korea. At the Potsdam Conference of 1945 the country was divided at the 38th parallel into a Soviet zone (North Korea) and a U.S. zone (South Korea). Attempts to reunify the zones into a single Republic of Korea failed. In the political and military upheaval that followed, millions of uprooted and homeless people wandered over the country. In 1955 the number of repatriates and refugees from Japan and North Korea was estimated at 4 million. These uprooted people had increased the South Korean population by about 25 per cent as compared with its size in 1945 (International Labor Office 1959, pp. 124-125).

The South Korean government's task of rehabilitating and reconstructing its country was immeasurably complicated by the presence of these destitute and homeless people. A comprehensive program of resettlement and integration was begun under the direction of the provincial governments with the aid of the UN Civil Assistance Command (later Korean Civil Assistance Command) and the Office of the UN Command Coordinator for Korea. The UN Korean Reconstruction Agency (UNKRA), which was created by the General Assembly in 1950 and terminated in 1959 and was supported by contributions from 34 UN members and five nonmember states, provided food, shelter, and medical treatment for hundreds of thousands of refugees and lent skilled staff to the UN Command for the control and distribution of relief. In August 1947, the UN reported that all but 3 million had been resettled or given new

homes. By mid-1961, 1.6 million of these had been resettled; a further 1.4 million were living in barely human conditions on government-provided land in the area of Seoul and Pusan, an estimated half million of them—the so-called "hard core" cases—supported by Protestant churches, Korean World Service, and by other voluntary agencies (the source of this information is the *Newsletter* of the U.S. Committee for Refugees, in its issues for June 1961 and July-August 1962; other sources put the total number of refugees from North Korea to South Korea during the period 1948-1953 at about 5 million).

**Vietnam and Laos.** Mass refugee movements occurred during and after the armistice conference, held in Geneva in July 1954, which terminated hostilities in Vietnam, Laos, and Cambodia. Article 1 of the convention provided for the division of Vietnam at the 17th parallel (a "provisional military demarcation line"): the Vietminh (People's Army of Vietnam) were to the North, and the French Union forces to the South. Civilians were to be allowed to move freely from one zone to the other, and the authorities of each zone were to assist their movement. By May 18, 1955, the final evacuation date, 860,000 persons (of whom 676,000 were Catholics) were known to have left North Vietnam for South Vietnam (Corley 1958/1959, p. 526); the actual total, however, may have been about 960,000. In addition to Vietnamese, about 42,000 seminomadic tribesmen joined the flight to the south (*ibid.*, p. 528). Because of communication problems and the opposition of Vietminh forces, their flight was a difficult and dangerous undertaking. French military and civilian planes flew many refugees south, and the U.S. Navy task force transported others by sea from embarkation points along the coast (Lindholm 1959, pp. 63-76). Two Vietnamese organizations, the Refugee Commission and the Catholic Committee on Resettlement of Refugees, were aided by the governments of the United States, France, the Philippines, New Zealand, and Australia, and also by UNICEF, WHO, and various private charitable organizations (Corley 1958/1959, p. 528). By 1960, 315 refugee villages had been established for the newcomers on undeveloped agricultural land and, in the normal course, the integration of the refugees would soon have been completed (Schechtman 1963, p. 168).

However, in face of gradually intensifying military and guerrilla activities in the south since 1954, large numbers of refugees have moved from unsafe areas in search of shelter and protection. These later movements have occurred for a va-

riety of reasons: panic flight from areas of military operation; escape from Vietcong terrorism, extortion, and recruitment; and movement away from communist-controlled areas, both at the urging of religious leaders and as a result of government resettlement programs. In addition, some have left their homes because of typhoons and floods. By January 31, 1966, the Agency for International Development (AID) reported an estimated 1,001,808 refugees in South Vietnam (U.S. Committee for Refugees 1966, p. 23); at about the same time, the Office of Refugees and Migration Affairs of the U.S. Department of State estimated that since 1961 1.4 million South Vietnamese had become refugees in South Vietnam. It seemed likely that the actual movements exceeded both these figures. Another estimate by the U.S. Department of State (based, like its other estimate, on hearings before a subcommittee of the Senate Committee on the Judiciary held during 1965 and 1966) put the number of Laotian refugees from areas of Laos under communist control at about 350,000 during the period 1960-1966.

The Government of the Republic of Vietnam, with the assistance of AID and some 24 American voluntary agencies, has developed an ambitious program of assistance for emergency and long-range services, including programs that attempt to increase the refugees' productivity in resettlement areas. It is hoped that as many of the refugees as possible will eventually return to their villages in peace (see also U.S. Congress, Senate 1965a; U.S. Congress, Senate 1966a).

**China.** Invading Japanese armies caused refugee movements from southern China as early as 1938, but the major flow of refugees resulted from the protracted Chinese civil war, the final establishment of a Chinese communist government in 1949, and the transfer of the Nationalist government to Formosa. Hundreds of thousands of Chinese fled in various directions. An estimated 337,000 fled to the east and southeast Asian countries, where they lived in precarious conditions (Schechtman 1963, pp. 310-322). By 1960, 14,000 of these had been assimilated in the northern part of Laos, being ethnically the same as the Laotians and sharing the same religion. The great majority found refuge in the British Crown Colony of Hong Kong, where, of the 1962 population of 3.2 million, more than 1 million were estimated to be refugees (Schechtman 1963, p. 312; compare United Nations, Hong Kong . . . 1955, p. 955). But the true figures are difficult to determine, for mingled with the refugees were many former Hong Kong residents returning after the war. It is also estimated

that, during the period 1948-1966, about 40,000 European refugees entered Hong Kong and the Philippines from mainland China.

In May 1962, in self-defense against this human tidal wave, the British blockaded the borders in an attempt to control and limit the inflow of refugees; the communists, too, who had hitherto done little to prevent the exodus, now took steps to end it. But the inflow continued by sea, and in 1962 an additional 200,000 refugees were added to Hong Kong's teeming millions. Although British policy in Hong Kong incurred much criticism, re-creation was tempered by awareness of the difficulties facing the authorities there. It had to be admitted, too, that the remaining countries of the free world had failed to offer an alternative place of refuge to most of the fugitives from Chinese communism. On June 4, 1962, President John F. Kennedy did indeed authorize the parole of specified Hong Kong Chinese, and by June 1966 over 14,000 of them had been paroled into the United States. Brazil, too, accepted a limited number. The offer of the Chinese Nationalist government on Formosa to resettle 50,000 refugees was viewed with skepticism, in view of that government's own food and population problems; only the Portuguese colony of Macao maintained an open-door policy. By early 1966 an estimated 80,000 Chinese refugees were in Macao, and approximately 1.25 million-2 million in Hong Kong (U.S. Committee for Refugees 1966). According to figures based on records kept by the British authorities in Hong Kong and by a joint committee consisting of representatives of the UN High Commissioner for Refugees, the International Committee for European Migration, and the U.S. Escape Program, about 2,080,000 Chinese refugees entered Hong Kong and Macao from mainland China during the period 1948-1966.

Since 1953 the Hong Kong government has been engaged in the building of multistory resettlement housing and the establishment of a vast network of small industries in an attempt to keep pace with the rapidly increasing population. With the aid of international and local voluntary organizations, they have been doing an exceptional job of integrating the mass of refugees, but these efforts have inevitably lagged behind the needs of the situation (U.S. Congress, House, Committee on the Judiciary 1966, pp. 11-13). Figures for those Chinese who went to Formosa are also hard to determine. Their integration was greatly aided by the Nationalist Chinese government and the Free China Relief Association (International Labor Office 1959, pp. 128-129).



**Tibet.** Communist China's assertion of authority over Tibet in 1950 and the Lhasa uprising in 1959 resulted in the flight of thousands of Tibetans over the Himalayas. There are about 43,000 Tibetan refugees in India, 7,000 in Nepal, some 3,000 in Bhutan, and 3,000 in Sikkim. The Indian government engaged in settling about 7,000 on land, while about 28,000 were resettled through their own efforts. The remaining 8,000 presumably found work of some kind. In Nepal, under a bilateral agreement between the governments of Nepal and Switzerland, the resettlement of these refugees is the responsibility of both the Swiss government and the Swiss Red Cross (International Council . . . 1965, p. 23). Emergency aid was provided by the U.S. government and by private groups. The International Committee for the Red Cross (ICRC) aided in the vocational training and resettlement of Tibetan refugees, with a view to making them self-supporting.

### Refugees in the Middle East

**Palestine Arab refugees.** The UN decision to partition Palestine was followed by the 1948-1949 Arab-Israeli conflict and the flight of an estimated 500,000 people from their homes in the area of fighting. Of these, the majority were Arabs, with a smattering of Armenians, Greeks, and non-Jewish nationals of other countries. By June 30, 1966, 1,317,000 were registered with the United Nations Relief and Works Agency for Palestine Refugees in the Near East (UNRWA): 707,000 in Jordan, 307,000 in the Gaza strip, 164,000 in Lebanon, and 140,000 in Syria (UNRWA 1966). From the outset, the compilation of accurate statistics on the refugees was made extremely difficult; for political reasons, and out of a desire for additional material aid, figures were inflated by false registrations and unreported deaths (Gabbay 1959, pp. 165-171). UNRWA was established by the UN in 1949 as a temporary organization. But since substantial progress in the reintegration of the refugees, either by repatriation or by resettlement, has not materialized, the mandate of the agency has several times been extended. From May 1, 1950, to December 31, 1965, UNRWA spent close to \$535 million, contributed to a large extent by the United States.

By June 30, 1965, 70 per cent of the refugees received basic dry rations and 40 per cent had been sheltered in 54 camps, the rest having found their own accommodations. More than 228,000 children were going to school, 168,000 of them in the 406 UNRWA-UNESCO schools. UNRWA operates 10 vocational or teacher training centers

and undertakes basic preventive and curative health care through 88 clinics and other facilities; in addition, more than 250,000 refugees, mainly children, benefit from UNRWA's programs of supplementary feeding and milk distribution (United Nations, Relief and Works Agency . . . 1966). With the cooperation of the Arab host governments and the international organizations, UNRWA attempts the dual task of providing relief and assisting refugees to become productive and self-supporting.

However, the problem of the Palestine Arab refugees is as much political as socioeconomic. Until a political solution is achieved, these refugees will continue to symbolize the instability that, in 1967, resulted in yet another Arab-Israeli war.

**Israel.** The persecution of the Jews by Nazi Germany was an important factor in speeding up the establishment of the Jewish national home first envisaged in the Balfour Declaration of 1917. Immigration of Jews into Palestine, which had continued both openly and clandestinely since the end of World War I, became legally unrestricted upon the establishment of the state of Israel in 1948. An ambitious resettlement program has been pursued by the government of Israel, involving a total of 1,209,282 immigrants by 1964 (Israel . . . *Statistical Abstract*, 1965), the majority being refugees from the countries of central and eastern Europe, north Africa, and the Middle East. The Jewish Agency for Israel, the Joint Distribution Committee, and the United Jewish Appeal, as well as the Intergovernmental Committee for European Migration and the United Nations High Commissioner for Refugees, assisted in this work.

### Refugees in Africa

The attainment of independence by more than 35 African countries during the period 1951-1966 has been accompanied by a complex displacement and uprooting of peoples, which in turn has resulted in severe problems. Deep-seated tribal and ethnic rivalries in many of the newly independent countries and restrictive conditions under continuing white minority rule have resulted in a flight to neighboring territories, adding to the long-existing tendency to economic migration. The estimates below of the number of people involved in these movements must be treated with caution. The task of obtaining accurate figures on refugees—never an easy one—is here made even more difficult by a number of factors: the poverty and relative lack of development in the countries of origin and asylum; the complexity of motivations leading to these movements; the variety of aids and attempted solutions, ranging from repatriation

(as in the case of the Algerians and the Congolese) to attempts at local integration, especially in the case of irredentist movements (as with the Somalis).

Since 1961, an estimated 650,000 persons have fled into adjacent territories. A partial breakdown of the figures for African refugees indicates the following movements.

(1) As of late 1966, more than 300,000 refugees had made their way to the Republic of the Congo: at least 250,000 came from Portuguese Angola, 25,000 from Rwanda, and 40,000 from the Sudan, as well as an unknown number from various other countries. Accordingly, the UNHCR helped to establish and maintain coordination between the Congolese authorities, the UN Organization in the Congo, the International Red Cross, and a number of voluntary agencies.

(2) By the end of 1966, Burundi had an estimated 78,000 refugees from Rwanda and the Congo.

(3) Also by the end of 1966, the Central Africa Republic harbored more than 6,000 Congolese and 25,000 Sudanese.

(4) In the fall of 1965, about 2,000 Burundis crossed into Rwanda.

(5) About 50,000 refugees from Portuguese Guinea moved into Senegal during 1964 and 1965. In 1966 these refugees were benefiting from a joint program of the Senegalese government and the United Nations, as well as from bilateral aid from the United States and France.

(6) According to estimates available early in 1967, there were about 30,000 refugees in Tanzania, including 12,000 from Mozambique and an equal number from Rwanda. Most of the remainder were Congolese.

(7) In December 1965, Zambia requested emergency aid for some 5,000 refugees. By the end of May 1966 more than 1,000 Angolese refugees were in Zambia.

(8) By mid-1966, the number of refugees in Uganda was estimated at 140,000 (70,000 from Rwanda, 45,000 from the Sudan, and 25,000 from the Congo). About one third of this number arrived during 1965-1966.

(9) An unknown number of refugees of various origins have sought asylum in other African countries, including Kenya, Chad, and Ethiopia.

The governments of all these countries of asylum have developed emergency and resettlement programs in cooperation with the UNHCR (for further information, see United Nations, Office . . . 1966a).

(10) Since March 1960, a number of fugitives

from South Africa's apartheid policies have sought asylum in neighboring countries, notably Tanzania; similar flights have occurred from South-West Africa and from Rhodesia, totaling an estimated 1,500 by 1966 (U.S. Congress, Senate 1966a, p. 98).

The above survey of African refugees does not include an estimated 1 million or more Eastern Nigerians who fled into their Eastern Region homeland following the September 1966 massacre of Easterners in the Northern Region. The Eastern Region government set up the Eastern Region Refugee Commission in Enugu, Nigeria, in an effort to cope with this influx. Since they were national refugees, they were ineligible for aid under international refugee aid programs, but the burden of their resettlement would seem too great for Nigeria to bear alone, whatever the outcome of the 1967 civil war.

### Refugees in the Western Hemisphere

Political refugees have been a common phenomenon throughout the history of the Western Hemisphere. Small in number, they have chiefly comprised exiled political leaders and people of means seeking temporary asylum in other countries. In recent decades, however, the refugee problem has undergone a fundamental change (Inter-American Commission . . . 1965, p. 65). Refugees now tend to be fugitives from political persecution, the majority with little or no means, and their flight has been the cause of recurrent strain between their countries of origin and the countries in which they seek asylum.

The number of refugees in this area is difficult to determine, for much of this movement is of a complex, and often clandestine, nature. During recent decades, many Haitians, especially those in the skilled and professional classes, are known to have taken refuge in the Dominican Republic, the United States, and several Caribbean countries. Bolivians fleeing their country have been mainly workers seeking to escape their government's persecution; the majority of Paraguayan refugees are also opponents of the present regime. There is scant information about those who fled the Dominican Republic after the fall of the Bosch government in 1963 and the revolution of 1965.

Cuba. The major movement of population in the Western Hemisphere, and one which has caused considerable concern to the Latin American countries and the United States, has been the mass exodus of Cubans since the revolution of 1959. By October 1963 they totaled an estimated 350,000. About 275,000 crossed to the United



States. The rest fled to Spain, Puerto Rico, and various Latin American countries. Many of these came under the protection of the UN High Commissioner for Refugees. The United States federal government, through the Department of Health, Education and Welfare, provided a financial assistance program, as well as educational, retraining, and vocational programs. It established a policy of resettling as many of the refugees as possible outside of Miami, the main entry center. A number of voluntary resettlement agencies were contracted with to furnish resettlement services on behalf of the government. By December 1, 1965, nearly 100,000 of the 185,000 refugees registered at the Cuban Refugee Center in Miami had been resettled to self-supporting opportunities in 3,000 communities throughout the 50 states and Puerto Rico. A further influx of Cuban refugees into the United States began in December 1965, following an agreement between the Cuban and United States governments for reunion of refugee families. Under this agreement about 4,000 Cubans per month were being airlifted into the United States. The total arrivals by December 9, 1966, were 50,051, of whom 76 per cent had been resettled. By December 1966 the federal government had spent more than \$200 million, not including the additional sums and services provided by the non-governmental agencies (Holborn 1965a). The majority of Cubans, on entering the United States, were granted no more than parole status, which greatly hindered their economic integration. But in October 1966 Congress authorized an adjustment of their status to that of permanent residents.

### Toward a solution of refugee problems

The League of Nations, in its attempt to solve the refugee problems created by World War I, was severely handicapped by a limited conception of its powers and responsibilities. It was assumed (1) that those problems were of a temporary nature and that a final solution could be achieved; (2) that the mandate of the League's organs included only refugees so designated by international arrangements; (3) that the responsibilities of its international organs were not operational; and (4) that embarrassment of League members could be avoided by eschewing actions which would favor one country at the expense of its political opponent (Warren 1958, Macartney 1934).

It had become obvious that the western European countries were in no position to offer further hospitality to refugees unless these countries could be relieved of earlier arrivals. The Intergovernmental Committee on Refugees (IGCR), created

in 1938 on the initiative of President Franklin D. Roosevelt in order to help refugees from Germany and Austria, therefore embarked on an international program for planned and assisted migration (Warren 1958, p. 112). This was a new dimension in the conception of an effective refugee program, and one which was carried over into the period after World War II. Thus the refugee became a migrant, and as such had to fit the immigration requirements of the country of resettlement.

**United Nations activities, 1944-1951.** During the gradual liberation of Europe by the Allied forces, the tasks of assistance to refugees and displaced persons from eastern Europe were divided between the High Commissioner for Refugees of the League of Nations, the IGCR, and the United Nations Relief and Rehabilitation Administration (UNRRA). The last, created under an agreement signed in Washington on December 9, 1944, by 44 nations, provided care and maintenance in camps for refugees and displaced persons in its operational areas and aided them in their repatriation (Woodbridge 1950). IGCR gave material assistance to those in other areas and prepared for resettlement those who could not be repatriated; the High Commissioner continued to provide legal and political protection. According to estimates in 1946, there were some 1,675,000 refugees for whom new homes had to be found (Holborn 1956, chapters 10, 11).

Even before its formal inauguration, the United Nations was confronted with the task of dealing with millions of uprooted people, and it has been continually harassed by the problem ever since. Like its predecessor, it recognized the refugee problem as one of immediate urgency and as international in scope and character. It also followed League precedent in regarding the enormous and widespread refugee and displaced persons problems created by World War II and its aftermath as a passing phenomenon for which lasting solutions could be found. Thus the organs established by the United Nations to deal with refugees have all been of a temporary nature.

In December 1946 the UN General Assembly approved the creation of the International Refugee Organization (IRO) to take over all responsibilities from IGCR and UNRRA in July of the following year. The Preparatory Commission of IRO (PCIRO) shouldered these responsibilities until the constitution of IRO came into force on August 20, 1948, delayed until this date by the need for ratification by a fifteenth state. IRO then embarked on a comprehensive large-scale program and worldwide operational field activities.

By December 1951, IRO had spent \$430 million contributed by its 18 member governments (the United States contributed \$237 million). In carrying out its tasks, IRO administered a network of camps; it provided housing, food, and medical care; it arranged for the rehabilitation and retraining of refugees, as well as for their legal protection; it negotiated agreements for resettlement, brought the refugees to ports of embarkation, and, in a vast shipping operation, transported them overseas in its own ships (Holborn 1956, chapter 21).

However, by 1950 there were still an estimated 1,250,000 or more refugees throughout the world. Of these, 400,000 were registered with IRO but had not yet been settled; instead, they remained in Germany, Austria, and Italy, with an unknown but substantial number of unregistered out-of-camp refugees. In addition, new refugees from the countries of eastern Europe continued to flow into western Europe and were in need of some form of international assistance when IRO was liquidated.

International activities since 1951. The Office of the United Nations High Commissioner for Refugees (UNHCR) and the Intergovernmental Committee for European Migration (ICEM) were created to help complete the unfinished task of resettling refugees. The Office of the UNHCR was established by the UN General Assembly under the Statute of the UNHCR in 1950. The office is an integral part of the United Nations, but with the independence necessary for the effective carrying out of its mandate. The High Commissioner promotes, organizes, coordinates, and supervises international action on behalf of refugees protected by former international agencies and on behalf of any person who is outside the country of his nationality (or if he has no nationality, the country of his former habitual residence) because he has been persecuted by reason of his race, religion, nationality, or political opinion, and cannot—or, owing to such fears, does not wish to—avail himself of the protection of that country. Once a refugee acquires the nationality of his country of residence or resumes his former nationality, he is no longer under the mandate of the High Commissioner (for a summary of these and other aspects of the international program, see Holborn 1965b).

ICEM was formed on the initiative of the United States at the Migration Conference at Brussels in December 1951 as an agency for planned and assisted migration for European national migrants and refugees. It is independent of the United Nations, operating under its own constitution and

directed and financed by thirty emigration, immigration, and "sympathizing" countries. It inherited the International Refugee Organization's shipping fleet and world-wide network of embarkation and reception facilities, as well as its invaluable administrative experience. By December 1966, ICEM had transported 1,464,630 Europeans to overseas countries, of whom 726,000 were refugees—including Hungarian refugees and refugees under the United States Escapee Program (Intergovernmental Committee for European Migration, *Annual Report* 1966).

Nongovernmental voluntary agencies. Traditionally, the basic responsibility for dealing with an influx of refugees rests with the authorities of the country of asylum. However, from the time the refugee problem first became apparent, the voluntary agencies have stimulated public concern and action. They have implemented the programs financed by governments and have provided the human link between the individual refugees and the governmental and international agencies. They have been largely instrumental in organizing the sponsorships of refugees wishing to emigrate and have been active in making local integration a success by helping refugees to adjust as members of new communities. In 1963 the International Council of Voluntary Agencies (ICVA), an organization in Geneva that incorporates more than eighty voluntary agencies, was awarded the Nansen Medal in recognition of its work.

Political and legal protection. The right of asylum, by which a state can accord hospitality and protection to political refugees and refuse to repatriate them, is widely accepted and practiced in international law. But a state is under no legal obligation to grant asylum: refugees cannot claim it as a right. The granting of asylum to refugees is frequently the cause of political tension between the country of asylum and the country of origin—tension which in turn affects the refugee. The provisions of international law that determine the status of an alien assume that he has a nationality. If he does not, he has no clearly defined rights, and it may often be impossible for administrative officials, no matter how well disposed, to assist him. It is therefore of great importance for all parties concerned to secure some regularization of the status of the refugee and to designate an international authority to act as his representative.

A series of conventions drawn up between 1922 and 1951 sought to narrow the legal no man's land of refugees. They made provision for the granting of identity and travel documents (such as the so-called Nansen passport), for civil status, employ-



ment, education, and social assistance (United Nations, Department of Social Affairs 1949). Two basic principles were established by these agreements: (1) that the personal status of the refugee (legal capacity, right to marriage and divorce, adoption, etc.) shall be governed by the law of his country of residence; (2) that the *principle* of reciprocity shall not be refused to a refugee in the absence of *de jure* reciprocity. However, these rights have been established only for specified categories of refugees and have been granted only by those states which have ratified the conventions or act as if they have ratified them. Legal protection, therefore, is limited in practice to certain groups, and even for these groups is assured only in certain countries. Thus, although political protection by an international agency is extended to all eligible refugees, the appropriate international organizations can aid refugees only insofar as international conventions or negotiations make it possible.

A general Convention Relating to the Status of Refugees was adopted by a plenipotentiary conference at Geneva in 1951. This document declared that the term "refugee" should apply to all those who had been so considered under the Arrangements of 12 May 1926 and 30 June 1928, the Conventions of 28 October 1933 and 10 February 1938, the Protocol of September 1939, or the Constitution of the IRO (for all of which see Holborn 1938), and to those persons who, because of events occurring before January 1, 1951, have a "well founded fear of being persecuted for reasons of race, religion, nationality, membership of a particular social group, or political opinion"; are outside their country of nationality; and cannot "or owing to such fear" are unwilling to, avail themselves "of the protection of that country" (United Nations, Office . . . 1966b, p. 15). The Convention, signed by 24 states, came into force on April 22, 1954, and by December 1966 had been ratified by 50 governments.

Under the influence of the Convention, the term "refugee" has become a new legal concept. It has gradually been accepted "that the refugee has a special status that sets him apart from the ordinary alien because he is without any country's diplomatic protection" (Read 1962, p. 49). In addition, it is no longer relevant from where the refugee has come: eligibility for refugee status is determined by the relationship of the refugee to events (Weis 1954). Some measures in favor of refugees have been taken by various international governmental bodies and also through special protocols applying to international conventions and special provisions inserted in general international in-

struments [for a different view of this definition of "refugee," see MIGRATION, article on SOCIAL ASPECTS].

In recognition of the universal nature of the refugee problem and its indefinite duration, a draft protocol to the 1951 Convention drawn up by the UNHCR in consultation with the governments concerned was signed by the president of the 21st UN General Assembly on January 31, 1967. When ratified by six countries, it will make the Convention universally applicable to refugees and will secure for new groups of refugees the same status as that enjoyed by those already covered by the Convention.

Material aid. "Rights" and "status" do not in themselves constitute solutions to the problem of refugees, but are merely conditions for solutions. By the end of 1951 an estimated 400,000 non-settled refugees (130,000 of them living in camps) were still under the mandate of UNHCR. All of these needed material assistance and either resettlement by migration to overseas countries or local integration in the country of first asylum. The Ford Foundation made available a \$3 million grant for operational service. Following this precedent, the UN General Assembly authorized the High Commissioner to set up a three-year aid program, to be financed through voluntary contributions from governmental and nongovernmental sources and to run from 1955 through 1958.

While the Office was engaged in this program, it was requested by the UN General Assembly on November 9, 1956, to act as a general coordinator for the activities of governments and voluntary agencies on behalf of the more than 200,000 Hungarians pouring into Austria and Yugoslavia. Following the Hungarian emergency, a resolution of the General Assembly in 1957 emphasized the need for greater flexibility to cope with emergencies. The High Commissioner was then authorized to establish an emergency fund, to be made up by repayments of loans granted to refugees and to be used under the general directives of the executive committee of the UNHCR programs. Approximately 150,000 refugees had benefited from these programs by December 31, 1965: over 96,500 resettled on a permanent basis through UNHCR programs and some 43,000 through other means.

From January 1952 to January 1966, the UNHCR assisted 3.8 million refugees with international protection and/or material aid. Some 1.6 million ceased to be refugees either by returning to their country of origin or by acquiring citizenship in the country of asylum or final resettlement. But in spite of this reduction, the number of refu-

gees was 2.25 million as of January 1, 1966 (U.S. Committee for Refugees 1966).

Expansion of UNHCR responsibilities. The concept of "good offices" empowers the High Commissioner to provide assistance, without a specific mandate, to countries which request it. In 1957 the UN General Assembly recognized as a matter of concern to the international community the problem of nearly one million Chinese refugees who had, from 1949 onward, flooded into Hong Kong from mainland China (United Nations, Hong Kong . . . 1955). The assembly did not place these refugees under the High Commissioner's mandate, but asked him to lend his good offices on their behalf. The concept was applied again to assist the large number of Algerians (estimated in 1959 at 200,000), the majority without means, who had fled into neighboring Tunisia and Morocco. Initial emergency relief had been given by the International Committee of the Red Cross. In May 1957 the Tunisian government appealed to the High Commissioner for assistance, and, acting under a UN General Assembly resolution of November 1958, he joined with the League of Red Cross Societies in a relief operation to aid refugees in both countries. This task was successfully concluded by July 31, 1962. In addition, the High Commissioner organized relief operations in 1958, financed through voluntary contributions, for those who had fled their homes in Egypt as a result of the Anglo-French-Israeli Suez venture in October 1956. Finally, a resolution of the UN General Assembly in the World Refugee Year 1959 extended the authority of the High Commissioner to lend his good offices for the welfare of nonmandate refugees. This resolution provided an avenue of assistance to the new nations of Africa beset by the problems of population movements that had followed in the wake of independence. Togo was the first to seek help in this way, at the beginning of 1962 (Read 1962, p. 45).

In the African context, bearing in mind the fragile nature of the social and administrative structures of the newly emerging nations, the Office of the UNHCR was faced with an entirely new set of problems, requiring new approaches and solutions. Here its first task was to assess the immediate needs (notably food and medical care) and to alleviate these through an allocation from the Emergency Fund. The Office's second task was to seek "operational partners" experienced in emergency relief work, such as the League of Red Cross Societies, and to obtain additional sources of material aid. The next stage is to promote integration on the spot as a permanent solution for the refugees, since voluntary repatriation is often

impossible, and resettlement even more so. Here again, UNHCR works closely with the international voluntary agencies, as well as with the governments of the asylum countries.

An important aspect of refugee aid in the less developed countries, such as those of Africa, is that tackling the problem of integrating refugees in the country of asylum results simultaneously in the provision of developmental aid to the indigenous population. Thus, for example, at Mugeru in eastern Burundi, UNHCR has taken the lead in creating a village for 25,000 Rwandese refugees, with all this entails in infrastructure and organization. In this work, the High Commissioner often calls upon the UN operational agencies experienced in development work, such as ILO, FAO, WHO, and UNESCO.

While regional organizations like the Organization of American States (OAS) and the Organization of African Unity (OAU) can contribute much to the alleviation of refugee problems, it is important that the Office of the United Nations High Commissioner for Refugees be continued. Fortunately, the international community's awareness of the close interrelation between the political, social, economic, and psychological aspects of refugee problems has culminated in the strengthening of the High Commissioner's coordinating role to the point at which any given refugee situation can be dealt with in its entirety. It appears increasingly likely that only the High Commissioner, as the representative of the United Nations, can embody the impartiality and prestige required for a humane and effective solution of these problems.

LOUISE W. HOLBORN

[Directly related are the entries HOMELESSNESS; MIGRATION. Other relevant material may be found in INTERNATIONAL CRIMES; INTERNATIONAL ORGANIZATION; and in the biography of KULISCHER.]

#### BIBLIOGRAPHY

- AMERICAN ACADEMY OF POLITICAL AND SOCIAL SCIENCE 1966 *The New Immigration*. Edited by Edward P. Hutchinson. Annals, Vol. 367. Philadelphia: The Academy.
- CORLEY, FRANCIS J. 1958/1959 *Vietnam Since Geneva*. Thought 33:515-568.
- EDDY, CHARLES B. 1931 *Greece and the Greek Refugees*. London: Allen & Unwin.
- GABBAY, RONY E. 1959 *A Political Study of the Arab-Jewish Conflict: The Arab Refugee Problem*. Geneva: Droz.
- HOLBORN, LOUISE W. 1938 *The Legal Status of Political Refugees, 1920-1938*. *American Journal of International Law* 32:680-703.
- HOLBORN, LOUISE W. 1939 *The League of Nations and the Refugee Problem*. *American Academy of Political and Social Science, Annals* 203:124-135.



- HOLBORN, LOUISE W. 1952 The United Nations and the Refugee Problem. *Yearbook of World Affairs* 6:124-148.
- HOLBORN, LOUISE W. 1956 *The International Refugee Organization, a Specialized Agency of the United Nations: Its History and Work, 1946-1952*. Oxford Univ. Press.
- HOLBORN, LOUISE W. 1961 Intergovernmental Partnership for Planned Migration. *Migration* 1, no. 2:5-18.
- HOLBORN, LOUISE W. 1965a The Cuban Refugee Problem. Report prepared for the U.S. Department of Health, Education and Welfare, Cuban Refugee Program. Unpublished manuscript.
- HOLBORN, LOUISE W. 1965b International Organizations for Migration of European Nationals and Refugees. *International Journal* 20:331-349.
- ICVA News. → Published since 1962 by the International Council of Voluntary Agencies.
- INTER-AMERICAN COMMISSION ON HUMAN RIGHTS 1965 *Draft Report on the Political Refugees in America*. Prepared by the Secretariat of the Commission. OEA/Ser. L/V/II.11, Doc. 7 Rev. Washington: Pan American Union.
- INTERGOVERNMENTAL COMMITTEE FOR EUROPEAN MIGRATION Annual Report of the Director to the Council. → Published in mimeographed form since 1952.
- INTERGOVERNMENTAL COMMITTEE FOR EUROPEAN MIGRATION 1956 *Constitution, Adopted at Venice, October 19, 1953*. U.S. Department of State, Publication No. 5987. Washington: Government Printing Office.
- INTERNATIONAL COUNCIL OF VOLUNTARY AGENCIES 1965 The Refugee Problem Since 1955, by Jacques Vernant. I.C.V.A. General Conference Document No. 20F. Unpublished manuscript, I.C.V.A., Geneva.
- INTERNATIONAL LABOR OFFICE 1959 *International Migration, 1945-1957*. Geneva: The Office.
- International Migration. → Published since 1961. Titled *Migration from 1961 to 1962*.
- ISRAEL, HALISHKAH HAMERKAZIT LESTATISTIKAH Statistical Abstract of Israel, 1965.
- JAHN, ERHARD 1966 Developments in Refugee Law in the Framework of Regional Organizations Outside Europe. A.W.R. Bulletin 4, no. 2:77-90.
- KULISCHER, EUGENE M. 1943 *The Displacement of Population in Europe. Studies and Reports, Series O, No. 8*. Montreal: International Labor Office.
- KULISCHER, EUGENE M. 1948 *Europe on the Move: War and Population Changes, 1917-1947*. New York: Columbia Univ. Press.
- LINDHOLM, RICHARD W. (editor) 1959 *Vietnam: The First Five Years*. East Lansing: Michigan State Univ. Press. → See especially pages 63-76, "The Role of the United States Navy."
- MACARTNEY, CARLILE A. 1934 Refugees. Volume 13, pages 200-205 in *Encyclopaedia of the Social Sciences*. New York: Macmillan.
- MCDONALD, J. R. 1965 The Repatriation of French Algerians, 1962-1963 *International Migration* 3:146-155.
- Migration News. → Published since 1952.
- ORGANIZATION OF AMERICAN STATES 1964 *Immigration in Latin America*. Washington: Pan American Union.
- PROUDFOOT, MALCOLM J. 1956 *European Refugees, 1939-1952: A Study in Forced Population Movement*. Evanston, Ill.: Northwestern Univ. Press.
- READ, JAMES M. (1951) 1953 *Magna Carta for Refugees*. Rev. ed. New York: United Nations, Department of Public Information.
- READ, JAMES M. 1962 *The United Nations and Refugees: Changing Concepts*. International Conciliation No. 537. New York: Carnegie Endowment for International Peace.
- REES, ELFAN (1957) 1959 *We Strangers and Afraid: The Refugee Story Today*. New York: Carnegie Endowment for International Peace.
- SCHECHELMANN, JOSEPH B. 1946 *European Population Transfers, 1939-1945*. Oxford Univ. Press.
- SCHECHELMANN, JOSEPH B. 1963 *The Refugee in the World: Displacement and Integration*. New York: Barnes.
- SIMPSON, JOHN H. 1939a *The Refugee Problem: Report of a Survey*. Oxford Univ. Press.
- SIMPSON, JOHN H. 1939b *Refugees: A Review of the Situation Since September 1938*. London and New York: Oxford Univ. Press.
- SMITH, RICHARD F. 1966 Refugees. *American Academy of Political and Social Science, Annals* 367:43-52.
- UNITED NATIONS, DEPARTMENT OF SOCIAL AFFAIRS 1949 *A Study of Statelessness*. Lake Success, N.Y.: United Nations.
- UNITED NATIONS, HONG KONG REFUGEES SURVEY MISSION 1955 *The Problem of Chinese Refugees in Hong Kong*. Report submitted by Eduard Hambro. Leiden (Netherlands): Sijthoff.
- UNITED NATIONS, KOREAN RECONSTRUCTION AGENCY Report of the Agent General. → Reports submitted to the sixth through the fourteenth sessions of the General Assembly.
- UNITED NATIONS, KOREAN RECONSTRUCTION AGENCY 1956 *UNKRA in Action*. New York: United Nations.
- UNITED NATIONS, OFFICE OF THE UNITED NATIONS HIGH COMMISSIONER FOR REFUGEES Review. → A monthly newsletter.
- UNITED NATIONS, OFFICE OF THE UNITED NATIONS HIGH COMMISSIONER FOR REFUGEES Statute of the Office of the United Nations High Commissioner for Refugees. → Published since 1951.
- UNITED NATIONS, OFFICE OF THE UNITED NATIONS HIGH COMMISSIONER FOR REFUGEES 1955 *A United Nations Plan for Refugees*. Geneva: United Nations High Commissioner for Refugees.
- UNITED NATIONS, OFFICE OF THE UNITED NATIONS HIGH COMMISSIONER FOR REFUGEES 1958 *The Convention of 28 July 1951 and the International Protection of Refugees*, by C. A. Pompe. Paper HCR/Inf/42. Geneva: United Nations High Commissioner for Refugees. → A press release.
- UNITED NATIONS, OFFICE OF THE UNITED NATIONS HIGH COMMISSIONER FOR REFUGEES 1961 *Forty Years of International Assistance to Refugees*. Geneva: United Nations High Commissioner for Refugees.
- UNITED NATIONS, OFFICE OF THE UNITED NATIONS HIGH COMMISSIONER FOR REFUGEES 1963 *The Red Cross and the Refugees*. Geneva: United Nations High Commissioner for Refugees.
- UNITED NATIONS, OFFICE OF THE UNITED NATIONS HIGH COMMISSIONER FOR REFUGEES 1966a *African Refugees: Report*. Geneva: United Nations High Commissioner for Refugees.
- UNITED NATIONS, OFFICE OF THE UNITED NATIONS HIGH COMMISSIONER FOR REFUGEES 1966b *Convention Relating to the Status of Refugees*. Geneva: United Nations High Commissioner for Refugees.
- UNITED NATIONS, RELIEF AND WORKS AGENCY FOR PALESTINE REFUGEES IN THE NEAR EAST Annual Reports of

- the Director. → Published since 1950. See especially page 20 in the 1966 issue.
- UNITED NATIONS, RELIEF AND WORKS AGENCY FOR PALESTINE REFUGEES IN THE NEAR EAST 1966 *UNRWA 1966 and the Palestine Refugees: Facts and Figures*. Beirut: United Nations.
- U.S. COMMITTEE FOR REFUGEES 1966 *World Refugee Report: Annual Survey Issue, 1966-1967*. New York: The Committee.
- U.S. CONGRESS, HOUSE, COMMITTEE ON THE JUDICIARY 1954 *Refugees in India and Pakistan*. Report prepared by Emanuel Celler. 83d Congress. Washington: Government Printing Office.
- U.S. CONGRESS, HOUSE, COMMITTEE ON THE JUDICIARY 1963 *Study of Population and Immigration Problems, Administrative Presentations (II), Refugees and International Migration*. Special Series, No. 10. Washington: Government Printing Office.
- U.S. CONGRESS, HOUSE, COMMITTEE ON THE JUDICIARY 1966 *Refugee Problems in Vietnam, India, and Hong Kong, British Crown Colony*. 89th Congress, 2d Session. House Report No. 1769. Washington: Government Printing Office.
- U.S. CONGRESS, SENATE, COMMITTEE ON THE JUDICIARY, SUBCOMMITTEE TO INVESTIGATE PROBLEMS CONNECTED WITH REFUGEES AND ESCAPEES 1965a *Refugee Problems in South Vietnam and Laos: Hearings*. 89th Congress, 1st Session. Washington: Government Printing Office.
- U.S. CONGRESS, SENATE, COMMITTEE ON THE JUDICIARY, SUBCOMMITTEE TO INVESTIGATE PROBLEMS CONNECTED WITH REFUGEES AND ESCAPEES 1965b *Refugees and Escapees*. 89th Congress, 1st Session. Senate Report No. 37. Washington: Government Printing Office.
- U.S. CONGRESS, SENATE, COMMITTEE ON THE JUDICIARY, SUBCOMMITTEE TO INVESTIGATE PROBLEMS CONNECTED WITH REFUGEES AND ESCAPEES 1965c *Refugees From Communism in Asia: A Study*. 89th Congress, 2d Session. Senate Report No. 59. Washington: Government Printing Office.
- U.S. CONGRESS, SENATE, COMMITTEE ON THE JUDICIARY, SUBCOMMITTEE TO INVESTIGATE PROBLEMS CONNECTED WITH REFUGEES AND ESCAPEES 1966a *U.S. Apparatus of Assistance to Refugees Throughout the World*. 89th Congress, 2d Session. Washington: Government Printing Office.
- U.S. CONGRESS, SENATE, COMMITTEE ON THE JUDICIARY, SUBCOMMITTEE TO INVESTIGATE PROBLEMS CONNECTED WITH REFUGEES AND ESCAPEES 1966b *Refugee Problems in South Vietnam*. 89th Congress, 2d Session. Senate Report No. 1058. Washington: Government Printing Office.
- U.S. CONGRESS, SENATE, COMMITTEE ON THE JUDICIARY, SUBCOMMITTEE TO INVESTIGATE PROBLEMS CONNECTED WITH REFUGEES AND ESCAPEES 1966c *Refugee Problems in Vietnam, India and Hong Kong, British Crown Colony*. 89th Congress, 2d Session. Report No. 1769. Washington: Government Printing Office.
- U.S. DEPARTMENT OF STATE, BUREAU OF SECURITY AND CONSULAR AFFAIRS 1966 *Annual Report of the Visa Office, 1966*. Washington: Government Printing Office.
- VERNANT, JACQUES 1953 *The Refugee in the Post-war World*. New Haven: Yale Univ. Press.
- WARREN, GEORGE L. 1958 *The Development of Inter-governmental Collaboration in Migration*. *Catholic Lawyer* 4:112-117, 161.
- WEIS, PAUL 1954 *The International Protection of Refugees*. *American Journal of International Law* 48:193-221.
- WEIS, PAUL 1960 *Le concept de réfugié en droit international*. *Journal du droit international* 87:929-1001.
- WEIS, PAUL 1966 *Territorial Asylum*. *Indian Journal of International Law* 6:173-194.
- WITKAMP, F. T. 1957 *The Refugee Problem in the Middle East*. *R.E.M.P. Bulletin* 5:3-51.
- WOODBIDGE, GEORGE 1950 *The Displaced Persons Operations*. Volume 2, pages 469-532 in *United Nations Relief and Rehabilitation Administration, UNRRA: The History of the United Nations Relief and Rehabilitation Administration*. New York: Columbia Univ. Press.

## II

## ADJUSTMENT AND ASSIMILATION

Refugees, or "displaced persons," are individuals who have involuntarily left their homes or communities and have been compelled by forces over which they have little or no control—war, invasion, persecution, or natural disaster—to change their place of residence, often from country to country. They are to be distinguished from voluntary migrants by the nature of the "push" factors which motivate them to move; however, many of their adjustment problems are similar. After World War II there were about twelve million displaced persons in Europe alone; the principal countries to which they moved were the United States, Canada, Australia, Israel, Argentina, Venezuela, the United Kingdom, and Brazil (Borrie 1959, p. 17).

The strain of prolonged displacement from their homes and communities, often combined with prolonged residence in temporary camps, serves to intensify the adjustment problems of refugees. The detachment and "normlessness," as well as the physical deprivation which accompanies this status, often results in what has been called "DP apathy," namely, an initial attitude of passivity and lack of initiative. This is also the result of dependency patterns that tend to develop in situations, such as in refugee camps, where all basic needs are met by the authorities. Observers have noted a certain detachment from the past and a feeling of lack of continuity resulting from constant interruption of stable patterns (Bakis 1955).

Moving from one social system to another requires that the refugee be resocialized into a new set of groups and norms. In the long run, adjustment implies loss of group visibility as members become indistinguishable from members of the host society. This is accomplished through a process of increasing conformity to the latter's norms and cultural patterns. The extent to which the group will in fact become invisible depends not only on members of the group itself but also on the extent



to which the host society demands complete conformity or tolerates differential cultural patterning. Once refugee groups have accepted certain over-all cultural patterns, there is usually an increasing acceptance of cultural pluralism on the part of the host countries [see ASSIMILATION].

### Criteria of adjustment

Defining the criteria of adjustment presents a particularly ambiguous problem when the "core culture," or culture of the host country, is unstructured or heterogeneous. This has been the case in Israel, where the host society is itself relatively new and is made up of large numbers of culturally diversified subgroups. The question then becomes to *what* the new group is supposed to conform. Even in countries like Australia, Canada, and the United States, the range of conformity of old-timers is itself wide. Furthermore, a newcomer tends to conform to the norms of the appropriate class or reference group with which he identifies, and the norms of these groups may differ among themselves within the social system of the host society. It is therefore sometimes unclear as to how close to the "core culture" a refugee must come before he can be considered "assimilated."

**Economic adjustment.** Successful entry of the major breadwinner into the occupational system of the host society is generally considered to be a major criterion of adjustment. It is also the first critical step taken into the new society. Most countries that have admitted large numbers of refugees, with the exception of Israel during its early years, have not suffered from a serious unemployment problem. Indeed, in recent years many of the countries that have accepted refugees have done so largely as a means of building up their manpower resources, although ideological considerations are, of course, always present. Getting a job upon arrival is therefore usually not a critical problem. Some countries select prospective immigrants in terms of their qualifications to fit into the labor demands of the society. However, it is worth noting that at times of economic recession it is generally the newcomers to the labor force on whom local workers will focus their hostility.

Entry into the economic system is often accompanied by considerable downward mobility. Having little choice, refugees are pressed to accept jobs at a lower status level than that to which they have been accustomed or for which they have been trained. Downward mobility produces frustration and tends to slow up other aspects of the adjustment process. Professional or intellectual refugees have been particularly frequent sufferers from this form of deprivation. Thus the fact of "having a

job" must be considered qualitatively in terms of its implications and symbolic meaning to the individual, if it is to be used in research as a criterion of adjustment.

**Social adjustment.** Satisfactory entry into the economic structure is not necessarily associated with successful adjustment in other spheres. Although the over-all satisfaction stemming from successful economic adjustment may in some cases be generalized to other areas, thus speeding the adjustment process, there are also cases of economic success coupled with a large measure of social isolation from the host society and retention of many traditional norms. In a sense the economic process of adjustment is the easiest and quickest; other processes of assimilation are generally slower.

There are at least two reasons for this differential patterning of the over-all adjustment process. Refugees may be employed in enterprises employing only other refugees. This is quite common during the early stages of settlement, particularly if the host authorities have provided temporary housing adjacent to employment opportunities. The work situation will then provide an opportunity for interaction only with other refugees, either from the same countries of origin or from other countries, but in any case not with members of the host society from whom local norms can be learned.

Furthermore, even if the refugee works side by side with members of the host society, the work situation may be limited to formal, instrumental relationships and may provide little overlap with the more informal spheres. The patterns that are learned will then be the ones associated essentially with job performance but not those having to do with other aspects of life in the new society. It is entirely feasible, even common, for the refugee to differentiate his behavior into an acculturated portion, which is associated with his work situation, and a traditional portion, in which he speaks his own language, maintains traditional social patterns, and socializes his children in the traditional manner.

Other criteria of adjustment that have been used are knowledge of the host country's language and culture, social relations with and membership in host groups, satisfaction with life in the host country, and identification with and conformity to the norms and patterns of the host country. Acceptance of the host country as one's "home," weakening of traditional group patterns, adoption of external symbols of membership in the host society in the form of clothing, food preferences, recreation, etc., are also usable criteria of adjustment.

Taft has noted that each of these general criteria may vary in terms of four dimensions. Does the

individual *want* to be assimilated? Does he *do* anything in this direction? Does he *perceive* that he is accepted or rejected? Is he *in fact* accepted by the host group? Criteria of adjustment must thus be viewed both from the point of view of the individual refugee or group itself and from the point of view of the host society; and the image of the process may differ from these two points of view (Taft 1963, p. 153).

### The process of adjustment

Two sides to the adjustment process must be considered: that of the refugees on the one hand, and that of the host society on the other. A satisfactory outcome of the adjustment process necessitates a certain complementarity of the two.

The degree of similarity of the refugee's traditional culture, as well as the general similarity in social structure of his former homeland to the host society, plays a major role in the adjustment process. In a general sense, the extent of such differences determines *how much* change will be required in the acculturation process. The social structure of the country of origin and the refugee's former place in it predetermine his skills and attitudes and thus direct him to a specific segment or stratum of the host society. Immigrants with few industrial skills or little experience with modern business procedures, for example, will have difficulty moving into these sectors of an industrialized Western society. Similarly, a differently structured class system in his country of origin will increase the newcomer's difficulty in moving into what he considers the appropriate stratum of the host society. This problem has been less severe in Western countries that have admitted refugees of European origin, and considerably more evident in a country like Israel, which has admitted large numbers of immigrants from Near Eastern and north African countries into a predominantly Western-oriented social system.

**Predisposition to change.** Inevitably a certain measure of change is demanded of the refugee in the course of his adjustment process. His own predisposition to change plays a major role in conditioning this process. Such a predisposition implies a willingness to accept new roles for himself, often abandoning traditional ones. It also implies a certain "time-perspective" that enables the refugee to see beyond present difficulties of adjustment to future goals. This will be accompanied by a relatively low level of resentment at the difficulties met with along the way (Zubrzycki 1956, p. 155).

On the whole, refugees are less predisposed to change than immigrants who have changed their country of residence voluntarily and who therefore

tend initially to identify more positively with the host society. In most cases refugees immigrate to the host country because it is the only, or one of the few, alternatives available to them. Their initial response may therefore continue to reflect the sort of "DP apathy" referred to above; his results in a relatively low predisposition to change. Another response is one of relief in finally escaping the uncertainties and insecurities of refugee existence and a determination to rebuild one's life; this may result in a relatively positive predisposition to change. The former is probably more characteristic of older refugees and the latter of younger ones.

The predisposition to change is also dependent on the extent to which the refugee perceives the host country as his permanent locus of settlement. In some cases refugees intend to use the host country only as a temporary refuge until they are able to return to their former country or to immigrate elsewhere; such an orientation results in a low predisposition to change. With the passage of time it often becomes clear that return to the country of origin will be impossible or undesirable; the pattern of orientation then tends to shift toward a more positive predisposition to change.

The perceived status differential between the culture of the host society and that of the refugee is important in conditioning his predisposition to change. When the host culture is viewed as superior to their own, refugees will be more motivated to abandon traditional norms and patterns and to adopt those of the host society. When the reverse is the case, and the refugee feels that his traditional patterns are superior, the predisposition to change will be lower.

Institutions that aim to maintain group solidarity among the refugees will function in accordance with the general predisposition of the group. For example, a native language newspaper or an ethnic church will attempt to speed entry into the host society in a group with a high predisposition to change but will try to reinforce group separateness in a group with a low predisposition to change.

**Level of expectation.** The expectations of the newcomers concerning the host society play a major role in their adjustment. When expectations are unrealistically high, frustration may result. Thus there is reason to believe that realistic knowledge of what to expect in the new society aids the adjustment process considerably. The point of reference to which newcomers compare the reality of the new society is often their country of origin; they tend to view it nostalgically, thus exaggerating its favorable features. The point of comparison is rarely the DP camp in which they may have lived immediately prior to immigrating, since that was



inevitably viewed as temporary, but rather the last stable community in which they lived (Shuval 1963, pp. 43-113).

Similarly, unrealistic expectations on the part of the host society impede satisfactory adjustment by refugees. Lack of knowledge of the newcomers' cultural patterns or expectations of too rapid conformity may lead members of the host group to feelings of hostility when their expectations are not met with quickly.

**Orientation of the host society.** The attitude of the host society toward the entry of refugees will be favorable when a labor shortage exists and the newcomers are seen as major manpower supplements; this has been the case in recent years in Canada and Australia. In these cases, economic expansion has tended to weaken the traditional objections of organized labor to immigration. In the United States, on the other hand, the labor unions have been instrumental in limiting immigration since the 1920s. Another factor inducing a favorable attitude on the part of the host society has been the ideological one, as in the case of Israel, where refugees have been welcomed unconditionally in terms of the over-all goal of that society to provide a homeland for all Jews.

With time, the over-all orientation of the host society often becomes differentiated with respect to various subgroups of refugees. Other factors having to do with color and cultural differences among subgroups of refugees may lead to patterned hostility on the part of certain members of the host society. Hostile attitudes toward groups already present in the host society may be generalized to specific appropriate groups of refugees (Shuval 1962).

**Visibility of the refugee group.** The sheer size of the refugee group plays a major role in the process of its adjustment. Size must be considered in relation to the size of the population of the host society or to that segment of the host society into which the refugees move. While the largest absolute number of refugees immigrated to the United States between 1946 and 1954 (1,700,000), they represented but a small proportion of the total population of the host country; the 790,000 who immigrated to Israel during the same period more than doubled the total population of the country. The size of the group is relevant because it affects its visibility and identifiability as a social unit. When the group is relatively small, its absorption can occur with few members of the host society being aware of or affected by the process. When it is proportionately large, absorption of the refugees turns into a major national problem, with virtually all members of the host society aware of and fre-

quently involved in one way or another in the process.

The social *homogeneity* of the refugee group also affects its visibility to members of the host society and consequently the rate at which it will successfully lose its group identity. When members of the refugee group are heterogeneous with respect to their class origins, religious identity, educational background, health status, or occupational potential, they are able to move more smoothly into the appropriate segments of the host society. If, on the other hand, they are highly concentrated in one such category (for example, lower class, Jews, unskilled workers, or chronically ill), the visibility of the group is increased and members of the host society continue for much longer to apply the appropriate label to all members of the group.

**Ecological patterns.** Ecological patterns of residence play a major role in determining patterns of adjustment. The initial tendency of immigrant groups—for reasons of in-group solidarity, government policy, or limitation of alternatives—is to locate themselves residually close to members of their own group. Such proximity tends to ease the process of adjustment for members of the group by lending support during the initial period of strain. At the same time, it tends to slow down acculturation in at least two ways. First, it reinforces traditional patterns of language and culture, thus slowing down conformity to norms and values of the host society. Furthermore, by maintaining group visibility, it focuses the attention of the host society on the group and on its nonconformity, often arousing hostility on the part of the members of the host group.

In Israel an attempt was made during the early years of refugee settlement to mix groups at random, but the many social problems that this generated led with time to a planned policy of ethnically homogeneous rural settlements. In terms of over-all adjustment, the impediment resulting from such ethnic isolation is probably overbalanced by the more favorable group adjustment and psychological stability resulting from it.

### Patterns of adjustment

The initial settlement of refugees often occurs in temporary camps. This has occurred, notably in Australia and Israel, because of the sudden large scale of the immigration and because the host society was typically unprepared, economically or socially, for the influx. In the United States and Canada, initial settlement has more frequently been on an individual basis, with assistance rendered through governmental or community organizations.

Camps tend to isolate residents from the general

stream of social life of the host society because of their physical isolation, as well as because of the necessary interaction of residents with other refugees, both socially and often in the work situation. The temporary nature of such arrangements tends to limit residents' ability to plan ahead or to structure their future lives in terms of more permanent goals. In this sense, camps in the host country tend to perpetuate the kind of normlessness and detachment that characterize the refugee prior to immigration. In addition, Borrie (1959) has noted that continued residence in a situation in which the individual is supplied with his basic necessities, while authority remains in the hands of the government, tends to increase the individual's expectations that he is entitled to certain rights and services. In an attempt to break this pattern of dependency, Israel abandoned the policy of complete support for immigrants during the first years of immigration by establishing "ma'abarot," that is, communities in which housing was provided, but residents were expected to support themselves.

During the initial stages of adjustment, refugees have been observed to follow two different patterns of adjustment: increasing conformity among those with a generally positive orientation toward the host society, and increasing withdrawal and lack of conformity to the norms of the host society by those groups who are initially disappointed or frustrated (Shuval 1963, pp. 117-138; Taft & Doczy 1962, pp. 64-67). Over a longer time span, a gradual pattern of conformity to norms of the host society appears among most refugees. On the other hand, there is reason to believe that certain fundamental aspects of the traditional culture are peculiarly tenacious; while many changes may occur, the ethnic group tends to retain its identity in certain spheres for a surprisingly long time (Glazer & Moynihan 1963).

The host society is often interested in focusing refugee settlement in rural or nonurban areas. In some countries this policy is based on considerations of economic planning and security. For refugees originating in urban communities, the over-all adjustment process is often made more difficult when the need to adapt to a nonurban community is superimposed on the general requirements of adjustment to a new society. Despite this policy, there seems to be a voluntary trend on the part of refugees to filter back to urban centers. Often this internal migration is motivated by the need to be near kin or near urban centers of employment that are perceived as more favorable.

It is often technically difficult for the authorities of the host society to keep families of refugees together at the time of settlement. This has im-

posed a particular strain on groups that have traditionally lived in extended family systems; but even for groups accustomed to a nuclear family pattern of living, the strain of adjustment is heightened by isolation from familiar primary groups of kin. There seems to be reason to believe that detachment or uprooting from earlier social networks in the country of origin increases the need for stable, supportive relationships, particularly during the early years of adjustment to the new society. Such relationships are most effectively provided by family or, failing this, by individuals from the refugee's own country of origin.

JUDITH T. SHUVAL

[Directly related is the entry MIGRATION. Other relevant material may be found in MINORITIES.]

#### BIBLIOGRAPHY

- BAKIS, EDWARD 1955 "D.P. Apathy." Pages 76-88 in Henry B. M. Murphy, *Flight and Resettlement*. Paris: UNESCO.
- BORRIE, WILFRID D. 1959 *The Cultural Integration of Immigrants*. Paris: UNESCO.
- EISENSTADT, SHMUEL N. (1954) 1955 *The Absorption of Immigrants: Comparative Study Based Mainly on the Jewish Community in Palestine and the State of Israel*. Glencoe, Ill.: Free Press.
- GLAZER, NATHAN; and MOYNIHAN, DANIEL P. 1963 *Beyond the Melting Pot: The Negroes, Puerto Ricans, Jews, Italians, and the Irish of New York City*. Cambridge, Mass.: M.I.T. Press.
- INTERNATIONAL SOCIOLOGICAL ASSOCIATION 1955 *The Positive Contribution by Immigrants*. Paris: UNESCO.
- KOSA, JOHN (1957) 1958 *Land of Choice*. Oxford Univ. Press.
- MURPHY, HENRY B. M. 1955 *Flight and Resettlement*. Paris: UNESCO.
- SHUVAL, JUDITH T. 1962 Emerging Patterns of Ethnic Strain in Israel. *Social Forces* 40:323-330.
- SHUVAL, JUDITH T. 1963 *Immigrants on the Threshold*. New York: Atherton.
- TAFT, RONALD 1963 Applied Social Psychology, Ecological Studies of Immigrant Assimilation, and Scientific Psychology. *Australian Journal of Psychology* 15:149-161.
- TAFT, RONALD; and DOCZY, A. GEDEON 1962 *The Assimilation of Intellectual Refugees in Western Australia*. The Hague: Research Group for European Migration Problems.
- ZUBERZYCKI, JERZY 1956 *Polish Immigrants in Britain: A Study of Adjustment*. The Hague: Nijhoff.

#### REGION

A region is a homogeneous area with physical and cultural characteristics distinct from those of neighboring areas. As part of a national domain a region is sufficiently unified to have a consciousness of its customs and ideals and thus possesses a sense of identity distinct from the rest of the coun-



try. The term "regionalism" properly represents the regional idea in action as an ideology, as a social movement, or as the theoretical basis for regional planning; it is also applied to the scientific task of delimiting and analyzing regions as entities lacking formal boundaries.

Official status is given to regions in the statistics of countries like the United States, where the national census groups states and counties in statistical and economic areas. In the international sphere, a region may consist of a group of national states possessing a common culture, common political interests, and often a formal organization. Thus the Scandinavian countries, the Benelux nations (Belgium, the Netherlands, and Luxembourg), and, in the Western Hemisphere, the Organization of American States (OAS) are true regional groupings (the OAS consists of North and South American states, with the exception of Canada).

In tracing the history of territorial integration, historians have placed little emphasis on ethnically or culturally homogeneous regions, tracing rather the cycle of conquest, aggregation, and empire building, and its antithesis of sectionalism, revolt, self-determination, and movements for national independence. In the Hellenistic world and during the Pax Romana larger cultural structures appeared—empires brought together by force. It was not until after the interregnum of feudalism that smaller political structures, in the form of modern nations, finally emerged from such empires. Regions thus appeared when national territories grew to large size, incorporated new domains, and began to adopt federal forms of government. In a highly centralized country like France, the origins of regionalism can be traced back to differences between historic provinces; it also represents a movement toward decentralization (Hintze 1934). In modern Italy and Germany unification laid the basis for regionalism—a trend accentuated by the increasing importance of economic differentials. Russia as the Union of Soviet Socialist Republics faces most of the problems of ethnic, economic, and sectional differences. In Great Britain regionalism verges on localism, for it goes beyond the traditional differences between England, Ireland, Scotland, and Wales, extending to the imponderable distinguishing characteristics of such subregions as Wessex, Sussex, and Yorkshire, which have been best portrayed in literature. In its extreme form, regionalism may result in separatism. Regions involved in such conflict are commonly referred to as "sections"; and sectionalism may culminate in self-determination, secession, and independent nationalism, as it did for

the Irish Free State. If the movement is crushed, the section may again assume the status of a constituent region, as did the South after the American Civil War.

As concepts utilized in social science, region and regionalism appear at some midpoint between the community and the nation. Whereas formal boundaries delimit the jurisdiction of political units, if basic regions are to be delimited, social scientists must concern themselves with social realities—demographic and economic. Regionalization is thus a heuristic device undertaken to advance analysis, planning, and administration; it is a common thread that runs through studies of regionalism by diverse disciplines. Regions can thus be regarded as building blocks making up the structure of the larger sociocultural area. The term "subregion" is usually applied to the next unit in descending order, but analysts also use such terms as tract, precinct, center, zone, district, and province.

Region and regionalism are topics common to all the social sciences, and contributions to theory and method in this area have come from both the physical and the social sciences, including biology. Certain evolutionary biologists, for example, hold views at variance with most psychologists about the man-land relationship. Any exegesis of regionalism is likely to prove a gloss on Aristotle's dictum that man is a political animal. To this is added the corollary that man is by nature a territorial animal. To agree, however, that all behavior patterns observed in man's political and societal structure are potential in human nature has not proved sufficient to satisfy the more ardent social Darwinists. The association of fighting and rage responses of insects, rodents, birds, and mammals (especially the higher apes) with the defense of territory against invaders of the same species has been demonstrated in a series of biological studies. Man, by species, evolution, and survival, it is contended, has fixed and retained these drives. In the present revival of conflict theory, a work such as Robert Ardrey's *The Territorial Imperative* (1966) represents a return to the instinct hypothesis—in this instance, a supposed compulsion to hold and defend territory. Citing recent biological studies, Ardrey bases his theory of the "territorial imperative" on the hypothesis that lower species have an instinct to defend definite and precise boundaries. He then extends this hypothesis to explain the foundation of the national state, patriotism, and the universal prevalence of war. Few psychologists or social scientists now welcome a return to the instinct hypothesis as an explanation equally ap-

plicable to the varied phenomena associated with localism, feudalism, nationalism, sectionalism, regionalism, and community spirit. To many the jump from animal instinct to human social organization is made by ignoring much of the work done by students of social structure.

Thus the areal approach to the analysis of society poses a stark reality underlying many of the interpretations made by social scientists. According to economic analysis, the primitive "sustentation area," anchored to the soil and water complex, set boundaries to social interaction and thus both aided in developing social types and, by forcing inbreeding, developed physical types among the inhabitants. In modern society, power, administration, social consensus, and policy determination appear inseparably linked to areas. Moreover, in the localization of resources, industry, and finance, and in the tie-in of lines of transportation to economic centers, the economic order parallels the political at all levels—from the international level to that of the component elements of states and provinces. If territorial groups did not exist, political organization would have to call them into being in order to function. The sociologist develops his specialty around social groups; these groups converge and are organized around interests—locality being a major interest. Society itself is seen as a group of groups. The territorial basis develops as a major social interest at any group level, and below the national level this interest can be seen as focusing on the section or the region. As other interests, such as those of social class or occupational group, gain importance, the regional interest may recede.

**Contributions of the geographers.** Geographers provided the initial impetus to regional studies. Rigorous and precise analyses of areas are to be found in the works of physical geographers, who are devoted to the integration of the findings of such specialties as climatology, physiography, soil science, and ecology. Geography set the mode of analysis early for differentiating world regions, as exemplified by the work of the French human geographers Paul Vidal de la Blache (1922) and Jean Brunhes. The point of view of the new cultural geography was well stated by Vidal de la Blache in 1903:

A region is a reservoir of energy whose origin lies in nature but whose development depends upon man. It is man, who, by molding the land to his own purposes, brings out its individuality. He establishes a connection between its separate features. He substitutes for the incoherent effect of local circumstances a systematic concurrence of forces. It is thus that a region defines

and differentiates itself and becomes as it were a medal struck off in the effigy of a people. (Lavissee 1903, p. 8)

It was natural that geography, with its emphasis on the analysis of land forms, should have proceeded from the natural to the cultural landscape. Dynamic assessment of the interaction of life forms and their habitats was first applied to biotic areas of indigenous plant and animal life and later to human-use regions seen as basically economic [see *ECOLOGY, article on HUMAN ECOLOGY*.] Areal analysis was thus held by some to be a function of the new ecology—plant, animal, and human. The full turn was made when culture and technology were given rein in the modern power complex, and, following G. P. Marsh, the regional phenomenon was viewed in terms of man's role in changing the face of the earth (International Symposium . . . 1956).

### Regional studies

Like the concept of space in philosophy, region as an idea general to all the social sciences has suffered from vagueness and thus from a too easy identification with any and all territorial units. The regional concept will contribute to the social sciences only if greater precision is introduced in its use. Recognizing this need, the Committee on Regionalism of the Association of American Geographers has distinguished between two types of regions—uniform and nodal. Uniform regions are homogeneous throughout, whereas nodal regions are homogeneous only with respect to internal structure or organization. The structure of the nodal region has a focal area, and the surrounding areas are tied to it by lines of circulation—communication and transportation (Whittlesey 1954, pp. 36–37). Uniform areas prevail when the physical environment is the important factor and agriculture is the prevailing mode of economy. Nodal regions appear with increased technology and the growth of large-scale transportation, wholesale distribution, finance, and manufacturing; in such regions metropolises perform the central work of the economy. An analysis of homogeneous regions and focal center areas, using the 67 metropolitan regions of the United States, showed almost no agreement between the boundaries of the two types of regions (Bogue 1957). Nevertheless, it seems clear from the same study that the uniform region and the nodal region concepts present equally valid and important ways of viewing the structure and processes prevailing in an inhabited territory.

The next problem in need of analysis concerns



the interrelations between these regional units. According to one viewpoint, the structure of national societies has developed as a complex of metropolitan regions, each consisting of its own constellation of communities. This view has been variously supported by, for example, N. S. B. Gras, an economic historian; Robert L. Dickerson, the British geographer; and R. D. McKenzie and Donald Bogue, American sociologists. As metropolitan areas expand, they may eventually form a continuous urban-strip region. This type of region—called the “megapolis”—has been analyzed by Gottmann (1961) in his study of the 43 contiguous metropolitan areas stretching from Boston, Massachusetts to Washington, D.C. Such an area makes new problems for the regional analyst. For example, the intense competition of closely placed centers points to the lack of regional unity. Moreover, it is difficult to assign hinterlands to these metropolises, when they are viewed as nodal regions. Because of the high population density of the megapolis, its components stand in need of regional-urban planning of a type as yet undeveloped.

Von Thünen's work on the isolated state (1826–1863) put location and spatial analysis in the forefront of accepted economic theory. Walter Christaller (1933) developed central place theory, a type of analysis which, strangely enough, had been undertaken by Charles Galpin in his work (1915) on rural communities in Walworth County, Wisconsin. The “place-work-folk” formulation of Le Play and the “valley section” of Victor Branford and Patrick Geddes brought social organization and advanced planning into the analysis of regional relationships. Culture area studies, carried forward under the leadership of Clark Wissler, utilized the location of physical and social traits to reconstruct regional cultures and their diffusion. Wissler's (1926) work on North American Indians was supplemented by Melville Herskovits' (Bascom & Herskovits 1959) culture area analysis of indigenous Africa. Regional studies, cultural in the best tradition, are also found in the work of the American geographer Carl O. Sauer and in Isaiah Bowman's (1931) analysis of “pioneer belts.”

It should be pointed out that the logic of regional study contravenes that of orthodox science: “Western science has been developed by specialization along lines of problem complexes or by abstraction and isolation of certain significant and meaningful ‘aspects’ of the chaotic reality. The social sciences are no exception to this principle. . . . The aim is always the establishment of gen-

eral principles, not the description of a single concrete society” (Heberle 1943, pp. 281–282). In regional analysis, phenomena are studied and related simply because they converge within a given area to affect the economy and culture of the particular societies. The regional survey, however, proceeds beyond a catalogue of traits into the analysis of an interacting system. Since causation cannot always be made explicit, such surveys often remain more descriptive than systematic—a fact which does not deny their usefulness.

Regional studies can be placed along an ascending scale of complexity, depending on the level of integration of variables. First, regions may be delimited on the basis of one variable; for example, the laying out of market areas and areas of wholesale distribution. Second, regions may be delimited on the basis of a complex of related variables, thus coming closer to cultural-areal reality. At the third level is the complex regional study that is developed in historical depth; it is calculated to show the functional unity of cultural, economic, and social traits of the region, as in W. P. Webb's *The Great Plains* (1931). Finally, interregional comparison furnishes an areal-cultural frame of reference for developing general theory about man, culture, or economy. [See CULTURE AREA.]

**Regionalism in the United States.** In the United States, regional studies have flourished under various auspices: regional science, area studies, American studies, literary regionalism, studies of administrative and planning regions, history, and sociology. Despite the fact that New England's regional economy and culture was an early development, it was not studied as a region, nor was New York, as long as both were accepted as the expression of the national pattern. When New England lost its hegemony, Seymour Harris wrote *The Economics of New England*. When the regional approach first appeared, it emanated from the West and the South, led in academic circles by Frederick Turner (1925) and Howard W. Odum (1909–1954; 1936).

Turner, a distinguished historian of the West, developed doctrines of the frontier, the section, and sectionalism, all of which undercut the continuity of the European cultural tradition in America, particularly the European concept of “undivided nationalism.” As the frontier matured, it became a distinct section, admitted to the Union but bent upon the pursuit of its own political and economic interests. Keenly aware of such developing differences, Turner viewed the United States as a sectional amalgam, something like a concert of European nations united in conflict, compro-

mise, and adjustment. States rights, he held, functioned as the constitutional shield for sectional demands.

Odum was a sociologist whose work was devoted to the problems of the South, including its lack of development and its strong sectionalism. As an alternative to sectionalism, he proposed a regionalism that would give precedence to national values. Odum thus advocated regional-national planning that would enable all areas to participate and benefit equally from the nation's resources, its capital wealth, and its social and political organization. He believed that the development of regional-national integration was necessary to the nation's survival.

Critics have pointed out certain limitations in regionalism, and it is true that, at its most superficial level, it approaches the method of the cataloguer. For example, the early geographers sought only to determine and list the geographic relationships and influences existing in the region; the criterion of a successful report was the number of traits catalogued. Moreover, the abandonment of the doctrine of determinism appeared to rule out the search for causation, and thus the cultural region and the physical region appeared juxtaposed but unrelated. In the early 1900s, economic and social differentials gave impetus to studies of regional equilibrium and integration, but as development furthered the trend toward economic equality, the obliteration of rural-urban differences, and the diffusion of a uniform culture, the basis for both regional analysis and regional ideology, appeared to diminish. Today, however, the problems of the newly independent, underdeveloped countries accent, as never before, the basic differentials in the status of world regions. The emerging problems inherent in international stratification highlight pressing dilemmas in population policy and regional-national planning (Lagos Matus 1963). Unfortunately, however, these problems are now concealed in the ideological conflict between communism and the Western world. During the era of expanding colonial empires, European scientists made excellent cultural and regional surveys of colonial areas preliminary to planning and development. Few such surveys are being made today; instead, reliance is placed on political dogmas. Thus supersensitive ex-colonials often expend aid funds to no appreciable economic advantage. The first of the new nations to return to the pragmatism of the older regionalism may prove the first to break out of this ideological impasse.

RUPERT B. VANCE

[Directly related are the entries CENTRAL PLACE; ENVIRONMENTALISM; REGIONAL SCIENCE; RURAL SOCIETY; SOCIAL DARWINISM. Other relevant material may be found in BUREAUCRACY; CITY, article on THE MODERN CITY; GEOGRAPHY; PLANNING, SOCIAL, article on REGIONAL AND URBAN PLANNING; TRANSPORTATION, article on SOCIAL ASPECTS; and in the biographies of BOWMAN; GEDDES; GRAS; LE PLAY; MARSH; ODUM; SAUER; THÜNEN; TROTTER; TURNER, VIDAL DE LA BLACHE; WISSLER.]

#### BIBLIOGRAPHY

- ARDREY, ROBERT 1966 *The Territorial Imperative: A Personal Inquiry Into the Animal Origins of Property and Nations*. New York: Atheneum.
- BASCOM, WILLIAM R.; and HERSKOVITS, MELVILLE J. (editors) (1959) 1962 *Continuity and Change in African Cultures*. Univ. of Chicago Press.
- BERTRAND, ALVIN L. 1952 Regional Sociology as a Special Discipline. *Social Forces* 31:132-136.
- BOGUE, DONALD J. 1949 *The Structure of the Metropolitan Community: A Study of Dominance and Subdominance*. Ann Arbor: Univ. of Michigan, Horace H. Rackham School of Graduate Studies.
- BOGUE, DONALD J. 1957 Nodal Versus Homogeneous Regions, and Statistical Techniques for Measuring the Influence of Each. *International Statistical Institute, Bulletin* 35, no. 4:377-392.
- BOGUE, DONALD J.; and BEALE, CALVIN L. 1961 *Economic Areas of the United States*. New York: Free Press.
- BOWMAN, ISAIAH 1931 *The Pioneer Fringe*. New York: American Geographical Society.
- CHRISTALLER, WALTER 1933 *Die zentralen Orte in Süd-deutschland: Eine ökonomisch-geographische Untersuchung über die Gesetzmässigkeit der Verbreitung und Entwicklung der Siedlungen mit städtischen Funktionen*. Jena (Germany): Fischer.
- DICKINSON, ROBERT E. (1947) 1956 *City Region and Regionalism: A Geographical Contribution to Human Ecology*. London: Routledge.
- DUNCAN, OTIS DUDLEY et al. 1960 *Metropolis and Region*. Baltimore: Johns Hopkins Press.
- GALPIN, CHARLES J. 1915 *The Social Anatomy of an Agricultural Community*. Research Bulletin No. 34. Madison: Univ. of Wisconsin, Agricultural Experiment Station.
- GALPIN, CHARLES J. 1918 *Rural Life*. New York: Century.
- GOTTMANN, JEAN (1949) 1955 *L'Amérique*. Paris: Hachette.
- GOTTMANN, JEAN (1961) 1964 *Megalopolis: The Urbanized Northeastern Seaboard of the United States*. Cambridge, Mass.: M.I.T. Press.
- HALL, ROBERT B. 1947 *Area Studies: With Special Reference to Their Implications for Research in the Social Sciences*. Pamphlet No. 3. New York: Social Science Research Council.
- HARRIS, SEYMOUR 1952 *The Economics of New England: Case Study of an Older Area*. Cambridge, Mass.: Harvard Univ. Press.
- HEBERLE, RUDOLF 1943 Regionalism: Some Critical Observations. *Social Forces* 21:280-286.
- HINTZE, HEDWIG 1934 Regionalism. Volume 13, pages 208-218 in *Encyclopaedia of the Social Sciences*. New York: Macmillan.



- INTERNATIONAL SYMPOSIUM ON MAN'S ROLE IN CHANGING THE FACE OF THE EARTH, PRINCETON, N.J., 1955  
1956 *Man's Role in Changing the Face of the Earth*. Edited by William L. Thomas et al. Univ. of Chicago Press.
- LACOS MATUS, GUSTAVO 1963 *International Stratification and Underdeveloped Countries*. Chapel Hill: Univ. of North Carolina Press.
- LAVISSE, ERNEST 1903 *Histoire de France depuis les origines jusqu'à la Révolution*. Volume 1, Part 1: Tableau de la géographie de la France, by Paul Vidal de la Blache. Paris: Hachette.
- LÖSCH, AUGUST (1940) 1954 *The Economics of Location*. New Haven: Yale Univ. Press. → First published as *Die räumliche Ordnung der Wirtschaft*.
- MEYER, J. R. 1963 *Regional Economics: A Survey*. *American Economic Review* 53:19-54.
- MUKERJEE, RADHAKAMAL 1926 *Regional Sociology*. New York: Century.
- ODUM, HOWARD W. (1909-1954) 1964 *Folk, Region, and Society: Selected Papers*. Chapel Hill: Univ. of North Carolina Press. → Organized to present Odum's sociological system. Contains a bibliography and a biographical sketch. Published posthumously.
- ODUM, HOWARD W. (1936) 1943 *Southern Regions of the United States*. Chapel Hill: Univ. of North Carolina Press.
- ODUM, HOWARD W.; and MOORE, HARRY E. 1938 *American Regionalism: A Cultural-Historical Approach to National Integration*. New York: Holt.
- STEWART, JULIAN H. 1950 *Area Research: Theory and Practice*. Bulletin No. 63. New York: Social Science Research Council.
- SUBBARAO, B. 1958 *Regions and Regionalism in India*. *Economic Weekly* 10:1215-1220.
- THÜNNEN, JOHANN H. VON (1826-1863) 1930 *Der isolierte Staat in Beziehung auf Landwirtschaft und Nationalökonomie*. 3 vols. Jena (Germany): Fischer.
- TURNER, FREDERICK JACKSON 1925 *The Significance of the Section in American History*. *Wisconsin Magazine of History* 8:255-280.
- VANCE, RUPERT B. (1932) 1935 *Human Geography of the South: A Study in Regional Resources and Human Adequacy*. 2d ed. Chapel Hill: Univ. of North Carolina Press.
- VANCE, RUPERT B. 1945 *All These People: The Nation's Human Resources in the South*. Chapel Hill: Univ. of North Carolina Press.
- VANCE, RUPERT B. 1951 *The Regional Concept as a Tool for Social Research*. Pages 119-140 in Wisconsin, University of, *Regionalism in America*. Edited by Merrill Jensen. Madison: Univ. of Wisconsin Press.
- VANCE, RUPERT B. 1960 *The Sociological Implications of Southern Regionalism*. *Journal of Southern History* 26:44-56.
- VANCE, RUPERT B.; and SMITH, SARA 1954 *Metropolitan Dominance and Integration*. Pages 114-134 in Rupert B. Vance and Nicholas J. Demerath (editors), *The Urban South*. Chapel Hill: Univ. of North Carolina Press.
- VIDAL DE LA BLACHE, PAUL (1922) 1926 *Principles of Human Geography*. New York: Holt. → First published posthumously as *Principes de géographie humaine*.
- WANDYCYZ, PIOTR S. 1957 *Regionalism and European Integration*. *World Affairs Quarterly* 28:229-259.
- WEBB, WALTER P. 1931 *The Great Plains*. Boston: Ginn.
- WHITTLESEY, DERWENT S. 1954 *The Regional Concept and the Regional Method*. Pages 19-68 in Preston E. James and Clarence F. Jones (editors), *American Geography: Inventory and Prospects*. Syracuse Univ. Press.
- WISCONSIN, UNIVERSITY OF 1951 *Regionalism in America*. Edited by Merrill Jensen. Madison: Univ. of Wisconsin Press. → Includes bibliographies.
- WISSLER, CLARK 1926 *The Relation of Nature to Man in Aboriginal America*. New York and London: Oxford Univ. Press.
- YALEM, RONALD J. 1962 *Regionalism and World Order*. *International Affairs* 38:460-471.

## REGIONAL PLANNING

See under PLANNING, SOCIAL. Also of relevance are REGION; REGIONAL SCIENCE; SPATIAL ECONOMICS.

## REGIONAL SCIENCE

Regional science, a new interdisciplinary field within the social sciences, focuses on the locational dimension of human activities in the context of their institutional structure and function and on the significance of this dimension in the understanding of social behavior and forms. The locational dimension identifies the spatial relations of people to their activities and also to the natural and man-transformed physical environment. Regional science relies heavily on mathematical models to frame its theories and draws on the theories and findings of other social sciences, particularly location theory (Alonso 1964; Isard 1956a, chapter 2). [See SPATIAL ECONOMICS, article on THE GENERAL EQUILIBRIUM APPROACH.]

The word "regional" implies the systematic approach to space in the sense of the human habitat. "Science" expresses the intention to apply rigorous techniques of investigation and to develop theoretical structures and concepts of general applicability. And "regional science" is meant to connote a field that transcends the bounds of any one social science discipline: it is related to regional economics, ecology, theoretical geography, regionalism in the sense of the political scientist, and parts of other social sciences, but differs from these fields in that it takes a more general approach to the role of space in social phenomena. Regional science also has a close affinity to a number of applied fields keyed to problems of adaptation to, or manipulation of, phenomena in space. Among these fields are city and regional planning, transportation, public administration, agronomy, and industrial engineering.

This article first discusses the formal structure

of regional science, including the types of analytic propositions and conceptual primitives that characterize its studies. We then turn to the research areas that have been most successfully explored and report on some of the major findings; we thereby illustrate the substantive content and the bounds of the field. Third, we comment on the relation of regional science to neighboring disciplines; and finally we touch on the educational programs in regional science and the state of the profession. We restrict this review to work in the United States; parallel activity is under way in Europe, in Japan, and, to a lesser degree, in India, Pakistan, Latin America, and elsewhere.

### Formal framework

Like other social sciences, regional science has three facets, which, in any one study, may emerge. Phenomena in space and choices related to such phenomena are presented and analyzed from each of three viewpoints: the normative, the descriptive, and the deductive. Thus, the distribution of urban places in a region, such as a nation or a part of a nation, can be approached *normatively*, as in the evaluation of conditions in many developing nations with high concentrations of population in metropolitan capitals (e.g., Harris 1959 and references cited therein); *descriptively* (or behavioristically), as in the analysis of existing distributions of hierarchies or rankings of urban centers (e.g., Zipf 1949, chapters 9 and 10); and *deductively*, as in the development of a theory of regional urban structure from propositions regarding personal demand functions, distances between consumers, economies of scale, and transport rates (e.g., Isard 1956a, chapter 11).

The regional scientist studies three major classes of decision makers: individuals (or households), entrepreneurs (businessmen or firms), and public bodies (such as city governments and regional planning organizations). These classes are of interest for two distinct reasons. First, each makes locational choices. Second, each responds to its environment, among other ways, by analyzing and anticipating the locational behavior and choices of the various elements of its environment. Thus, the regional scientist is concerned with developing appropriate techniques for studying locational choices as well as the consequences of such choices (for example, the resulting patterns of urban settlements).

The substance of the decisions represents yet another facet of the regional scientist's interest. He seeks to include in his analysis diverse economic, social, political, and other variables and

constraints as these are formulated in a spatial context (Isard & Isard 1965; Isard & Tung 1964). The decisions that are at the center of regional science analysis are those which establish the locations of various classes of activities in the environment. These decisions determine the scale of each class of activity at a given point in space and the nature, magnitude, and direction of flow of people, goods, funds, messages, and so forth between various loci of activity. The basic task, then, of regional science analysis is the identification of (a) location, (b) activity magnitude at the location, and (c) flow between locations. Regional science sees these as part of an interrelated system. Different locations involve the substitution of, for example, transportation charges for other factor costs. Different locations may also involve changes in the scale of operation as, for instance, market areas change. Furthermore, any change in location or scale at a given point has a concomitant effect on the flows emanating from and coming to the point. In short, regional science location analyses focus on questions of substitution of various categories of costs and revenues. The fundamental research problems are the determination of actual changes in accounts as spatial shifts occur, the identification of optimum location patterns under particular sets of welfare assumptions, and the identification of equilibrium conditions under postulated socioeconomic environments (Isard 1956a, chapters 4-10).

**Spatial structure.** The unique quality of regional science is not only its concern with the location decision and the individual decision maker within an environment but also its concern with the location itself and the locational framework. The locational framework, as a space, and locations, as points within the space, remain concepts that are analytic primitives. Points define locations in a bounded space; this space is the set of possible locations; and specification of points in the space establishes distances between points. But distances can be expressed in many ways. The distances of interest may be expressed in terms of physical units (e.g., miles), in terms of travel time, in terms of travel cost, or in terms of any of a variety of sociological attenuations. These distances can, to some extent, be translated into each other. Development of a satisfactory aggregative weighted distance measure awaits further progress of the field, yet it is clearly a requisite for a valid regional science (Dacey 1963; Deutsch & Isard 1961; Olsson 1965; Webber 1964).

**Regionalization.** Total space is not undifferentiated or homogeneous; neither are patterns of



activity spatially chaotic. Regionalization is the task of dividing the total set of points into relevant locationally defined subsets. In order to divide an area into regions, the analyst is faced with four tasks. He must identify the population, or what might be labeled the "constituent set," which is the set of points, lines, or areas (or the phenomena these represent) that are to be disaggregated. Second, the analyst must position these in the set of locations in space. Third, having a problem or a hypothesis in mind, he must establish one or more scales that "measure" a member of the population in terms of relevant attributes. Thus, the population is specified in terms of location and in terms of characteristics. Finally, cutoff points must be chosen along the attribute scales; this procedure places the member of a constituent set in one category or another, each category corresponding to a partitioning of the location set (Teitz 1962; 1964). The result is a pattern of regionalization; but because of the choices made at each of the four stages, an infinity of regionalization systems can be generated for any one constituent set. Little appears to be gained from the quest for an ideal set of regions; rather, each task or problem has associated with it a given set of regions in a system of regions (Isard 1956b).

A system of regions can be described in terms of several properties. First, is the system exhaustive? Second, is it disjoint? These two properties establish whether a given location is defined in terms of the region system and, if so, whether it is uniquely defined. Third, are the locations assigned to regions contiguous? Fourth, what is the principle of internal organization within a region? Thus, a disjoint regionalization is found in the designation of the Standard Statistical Metropolitan Areas (which are not exhaustive, however, with respect to the total territory of the United States). Coverage by radio stations is exhaustive (though clearly not disjoint). The counties within a state are both disjoint and exhaustive. Each of the states is made up of contiguous subregions, while the system of states, or the United States, is not, because of Alaska and Hawaii (Bunge 1962; Teitz 1964; Tobler 1963; Whittlesey 1954).

There are at least two basic systems of internal regional organization. One system of defining a region is by partitioning the population into spatially designated subsets that are internally homogeneous and externally heterogeneous. This gives rise to the identification of uniform regions. Examples of this approach are the basic soil and climate regions of the United States and the land-use mappings of urban areas. Each delineated area

has common internal characteristics that distinguish it from other areas. It should be added that the presumption is that socioeconomic variables of some consequence reflect this regional designation and partition. Further, such a region is an area where policies may be uniformly applied—a matter of considerable administrative interest. An alternative system of regionalization stresses internal cohesion or connection. This approach gives rise to the nodal-tributary region. It has been suggested that any one of numerous linkages or flows between populations specified by their location may be the basis for such a region.

The second regionalization principle is, perhaps, most clearly exemplified by the notion of a market area with characteristic flows to and from a central place. The set of consumers is spatially partitioned by the location of their suppliers. A set of cores and associated tributary areas results. The region is held together by the ties of interchange between purchaser and seller, and the regional boundary is the break point between the trading areas. At the same time, the set of nodal-tributary regions can form the framework for an administrative hierarchy. Here, too, there enters an element of homogeneity in the set of flow links that portray trade relationships (Berry & Garrison 1958; Berry & Pred 1961; Friedmann 1956; Garnsey 1956; Nystuen & Dacey 1961). [See CENTRAL PLACE.]

*Social physics.* Regionalization is not the only approach to the problem of classification of phenomena over space. One alternative has been the development of descriptive models that are analogues of the Newtonian physics of masses. Thus, it is possible to describe a distribution in terms of potential of the particular mass (e.g., in terms of people or of income earned). Such a potential measure represents the total influence at each point in space of all mass components, as the influence is attenuated by distance. The values of the potentials at each location indicate each location's proximity to other locations. This information is usually presented in map form. The usefulness of potential measures is that a number of socioeconomic variables have been found to correlate with the level of potential. An analytically related index of spatial structure is the population energy: the expected degree of spatial interaction between the masses at a pair of points is inversely proportional to the distance and directly proportional to the product of the population or other relevant masses at those locations. Again, certain flows have been found to behave according to this relationship. Socioeconomic analogues to force and gradient have also been developed. This social physics for-

mulation is highly suggestive; however, it remains mainly a descriptive and predictive tool. The measures and relationships it emphasizes are too aggregative to reflect the individual elements (Isard et al. 1960, chapter 11).

*Spatial statistics.* In addition to efforts directed to construction of macrogeographic measures of spatial structure, there have recently been successes in development of a formal spatial statistics. This work has concentrated on formulating measures of centrality, dispersion, and correlation over space analogous to those in traditional statistics. Corresponding to the mean, median, mode, and geometric mean there are spatial central measures. Formulas for standard distance deviations have also been derived. These measures have been calculated for a number of points in time to show spatial shift of the population (Dacey & Tung 1962; Neft 1965; Warntz & Neft 1960). Order statistics represents a further development in the analysis of spatial patterns. Both quadrant and nearest neighbor techniques have been applied. The use of these techniques has made possible tests for nonrandomness that provide the analyst with a test for similarity between patterns (Dacey 1963). [See GEOGRAPHY, article on STATISTICAL GEOGRAPHY.]

*Normative aspects.* The process of regionalization and the analysis of patterns of social phenomena have a purpose beyond construction of deductive models or precise description. A central purpose of regional science is to identify and analyze the problems of regions and to suggest solutions. Many of the regional science techniques have been adapted and developed for the purpose of making and implementing normative decisions by, for example, regional and metropolitan planning bodies.

The regional scientist characteristically attacks a region's problems with a comprehensive or systems approach, which leads him to an interdisciplinary view of the individual, entrepreneur, and public body. Traditional economic reasoning and economic costing can neither adequately explain nor optimally guide location decisions. Considerable thought has gone into modeling decisions within a framework in which entrepreneurial profit or individual utility maximization plays a less dominant role than in classical economics. Specifically, social profit, public welfare, and noneconomic considerations tend to play a major role in model formulation.

The systems approach also emphasizes social and economic interdependences in which the behavior of one actor is seen to affect all other

actors in the region and in which each actor has a characteristic and known relationship with others inside and outside the region. The concept of interdependence stresses not only the economic interrelations of men, firms, and organizations over space, but also their social and political interactions. Inquiry into these interdependences over space has proceeded at the highly abstract level of general equilibrium models as well as at the more operational levels of interregional input-output analyses and systems of interregional social accounts. Such operational models not only describe the structure of a system of regions but are particularly well suited to test and assess impacts of alternative public decisions and plans (Isard et al. 1960).

### Research techniques and findings

In discussing the major research in regional science, we shall emphasize operational techniques, particularly those which allocate scarce resources, given available technology, to the production of wealth for individuals and organizations motivated primarily by the wish to maximize utility or profit. These techniques can also be applied by public bodies concerned with social welfare, present and future. The range of techniques may be suggested by reference once more to the three classes of actors or decision makers: entrepreneurs, individuals, and governments.

*Entrepreneurs.* In accord with traditional economics, the entrepreneur is often seen as a decision maker who wants to maximize revenue less costs, where elements of both revenue and cost vary over space [see FIRM, THEORY OF THE]. These calculations are, it is assumed, made for a known and fixed time period over which conditions do not change. The entrepreneur possesses complete information, and, in the formal model of the location decisions, he knows how to use it. The location decision essentially involves a calculated substitution among transportation and other factor costs (Hoover 1948, chapter 3; Isard 1956a, pp. 222-235).

For example, one may wish to predict the decision of an entrepreneur selecting a location for a new manufacturing plant. For this prediction, comparative cost analysis may be employed, on the assumption that this analysis is a reasonable representation of the entrepreneur's decision process. The comparative cost approach consists of listing the locationally variable costs in serving a given market from a number of possible production sites and then picking the minimum cost point. A relatively simple extension of this model allows for



variation in revenues as alternative production sites under consideration tap different markets. A further extension allows for the effects of finished product transport costs on the effective demand of consumers. The iron and steel, aluminum, shoe and leather, and synthetic fiber industries are among those whose locational structure has been studied intensively in this manner. A review of the comparative cost technique will be found in Isard et al. (1960), which includes an extensive bibliography of pre-1960 studies.

The costs of operation of any given activity may be closely tied to the scale and costs of operation of other spatially and technically related activities. Recognition of the significance of external economies due to spatial juxtaposition has led to the development of the industrial complex technique. Plausible complexes of industrial units are examined, using input-output analysis and the substitution principles of comparative cost analysis [see INPUT-OUTPUT ANALYSIS]. A study of proposed oil-refining, petrochemical-fertilizer, synthetic-fiber complexes for the industrial development of Puerto Rico represents one example. Results of such research include identification of activity levels and locations optimal for the set of complexes studied. Industrial complex analysis can be used both to predict behavior of firms and to suggest policies to encourage industrial development. (For a discussion of this approach, together with a case study, see Isard, Schooler, & Viorisz 1959.)

Linear programming is one step beyond comparative cost and industrial complex analyses, which are restricted in the sense that they require preselection of output and shipment alternatives, which may well precondition results obtained. Linear programming, on the other hand, treats both the outputs at each location and shipments between locations as variables to be determined [see PROGRAMMING]. The classical transportation model of linear programming has had numerous applications. Further developments in linear programming allow determination of optimal production and shipment patterns for multilocation and multistage firms. Finally, more general models of whole industries emphasizing the use of locationally specified resources, including land itself, have been made operational. One interesting output of such studies is the generation of a set of location rents consistent with the optimal pattern of production shipments. These location rents provide a link between the linear programming models and classical location theory. Applications of the linear programming approach include Heady and Egbert

(1962), Henderson (1957; 1958), Manne (1958), Marschak (1958), Miller (1963), Stevens (1961), and Viorisz (1956).

The optimizing models that have just been described do not correctly depict the decision processes of entrepreneurs. The entrepreneur operates under constraints of imperfect knowledge; he also reacts to a host of noneconomic pressures and incentives. His calculus clearly extends to include elements of cost and revenue that the above formal models ignore; indeed, certain personal and institutional factors would appear to dominate in many location decisions. The divergence between optimal and actual behavior may also in part be explained by a secular trend toward homogeneity of locations in the United States and diminution of measurable differentials among locations. An increasing number of industries can be labeled "footloose," and decisions about their location are made in response to institutional and other factors rather than strictly economic considerations. Economically nonoptimal behavior is not as severely penalized. Studies by McLaughlin and Robock (1949), Hoover and Vernon (1959), Perloff and his colleagues (1960), and Fuchs (1962) analyze the noneconomic factors influencing locational decisions.

**Individuals.** Studies of household location decisions have focused on aggregates of individuals rather than on the location decisions of individuals. In contrast with the framework for studying entrepreneurial spatial decisions, where the unit of analysis is the individual firm, household behavior studies have been largely unable to identify any measurable criterion for optimal behavior or, for that matter, to explain manifested behavior at the level of the single household. One possible justification for dealing with individuals en masse is that individual behavior is assumed to be unpredictable in much the same way as the behavior of individual molecules. This assumption has led to development of the above-mentioned gravity and potential models of social physics and also to research studies of commuters, migrants, home purchasers, and shoppers. Such studies have typically analyzed behavior as a function of three factors: (1) distance, (2) socioeconomic characteristics of nodes, and (3) socioeconomic characteristics of the landscape between nodes. Thus, there has been a wide variety of studies verifying the hypothesis that interaction between a node and its tributary region (or between two nodes) decays with increasing distance. Further, the effective or perceived distance has been found to differ from physical distance, money-cost distance, and time distance because of the existence of intervening places, information flow

patterns, and institutional effects. (See, for example, Carrothers 1956; Marble & Nystuen 1963; Stouffer 1960; Warntz 1959; and Wingo 1961.)

The deductive models of mass behavior have either implicitly or explicitly been based on the assumption that interaction decreases with distance and increases with the number of opportunities. This, however, has led to two classes of similar models: first, those in which the physical analogy has been stressed, and, second, those in which the empirical utility of the models has been emphasized.

Certain normative approaches to mass behavior have been developed. Despite our previous statement that models of human behavior over space have been almost entirely oriented to mass probabilistic behavior, some attempts can be noted that ascribe optimizing behavior to the individual. A classic example is the so-called "traveling salesman" problem, in which the individual wishes to pick the shortest route that passes through each of a set of specific nodes (Dacey 1960). An example of the attempt to simulate household location decisions within an urban area has been developed in connection with a large-scale transportation study: each household is perceived to attempt to maximize its location advantage while competing with all other households for the available residential space (Herbert & Stevens 1960). This model is related to an abstract model in which households attempt to maximize their utility in the competition for urban space while landowners attempt to maximize returns from the land (Alonso 1960).

**Governments.** One of the characteristic problems faced by a government unit (say a city government, a planning organization for a metropolitan region, or a national development agency) is the anticipation, assessment, and adaptation to change brought on by major exogenous or outside impacts, for example, a reduction in military expenditures or an increase in population. Regional science techniques, particularly those based on recent developments in applied economics, are well suited to these tasks; application of the techniques by planning bodies is discussed at the theoretical level by Mitchell (1961), Davidoff and Reiner (1962), and Isard and Reiner (1962).

Such problems are best analyzed within a framework, or model, that relates the system's (e.g., nation's) demographic and economic variables to the regional demographic and economic variables of interest. The framework emphasizes the interdependence of the regions, and in practice it is constructed by synthesizing studies of various aspects of the system. Isard and his colleagues (1960, chapter 12) present in some detail various "chan-

nels of synthesis" and assess their operational significance.

The central feature of a typical framework is an interregional input-output matrix, which generates employment, population, migration, output, etc., by region when fed information specifying, by region, the final demands for each output. The basic findings of comparative cost and industrial complex studies can be incorporated into the framework by removing from the input-output matrix those industries or parts of industries whose location patterns are most meaningfully analyzed by use of these techniques. The results of these studies indicate specifically how certain industries and parts of industries are to be treated in specifying the final demands.

But comparative cost and industrial complex analyses require a set of basic assumptions for the system regarding, for example, birth rates, death rates, net immigration, technology, and tastes. From these assumptions, population, labor force, and average productivity of the system are projected for the key future year or years, and the social accounts (gross system product and gross system income) are then estimated. The social accounts are expressed in terms of levels of major aggregates, such as government expenditures and capital formation. System outputs by industrial sector are derived via input-output or other techniques. Initial regional markets are then established on the basis of current data and with the aid of relative growth charts, trend analysis, consumption expenditure studies, and ratios such as the coefficient of localization and location quotient (*ibid.*, chapter 7). The final demands by region are then fed into the fused comparative cost-industrial complex-interregional input-output framework to obtain the values of the regional variables. However, because of discrepancies, especially between estimates of initial regional markets and the set of regional markets consistent with the results of the computation, it may be necessary to re-estimate initial regional markets and rerun the framework. The rerun process is continued until results are obtained that are in harmony with the regional market assumptions underlying the results and with resource scarcities and community attitudes.

In practice, studies involving the synthesis of several regional science studies are often restricted. Some pertain to a single region rather than to a system of regions; some involve fewer techniques and yield less disaggregated results. On the other hand, recent developments in general social, political, and economic equilibrium analysis for a set of regions, and in cooperative procedures for a set



of regions, point to the possibility of still more extensive and comprehensive channels. Such systems are premised on a market containing social and political commodities as well as economic goods. Individuals function in a multiplicity of both active and passive roles; organizations engage in the production of economic and noneconomic goods, being motivated to maximize effective profits. Government units engage in production and distribution of programs of goods and services designed to maximize constituency welfare. Social norms and values are introduced and help govern actions chosen by the several behaving units. This broad social theory is designed not only to operate under the conventional price and market systems of economics but also with cooperative procedures developed from a fusion of game theory and normative concepts. The application of these cooperative procedures has been extended particularly to multi-region systems wherein elements of both harmony and conflict underlie development and investment programs (see Isard & Isard 1965; Isard & Smith 1966a; 1966b; 1967). Because of their ability to make consistent projections for a wide variety of variables of direct concern to public bodies, these channels of synthesis represent a direction of basic interest to central and regional planning agencies.

### Relation to neighboring disciplines

The relation of regional science to economics is close, since many of the economic variables (income, employment, prices), techniques (input-output, cost analysis, social accounts), and assumptions (maximizing behavior, efficiency of production) have been absorbed. The regional scientist spells out the regional structure of an economy and at the same time seeks to add to traditional economic analysis the locational and spatial dimensions of reality. Regional science also criticizes economics for failing to recognize the role of governmental authority and administration in the shaping of regional economies. Finally, the regional scientist views the national economy itself as the sum of a set of regional economies. (An extensive review of regional economics, including a survey of regional science and its relation to economics, may be found in Meyer 1963; see also Fisher 1957.)

Regional science has in common with political science a concern with the proper areal distribution of power, the number of levels and the structure of power in a hierarchy of governmental levels in which metropolitan government is of special interest, and the varying spatial pattern of the cost of administration. It also shares an interest in the new frontiers in political science in the search

for quantitative models that can be empirically established (Isard 1957).

Historically, the relation between regional science and sociology has been close, since many of those in the field have had strong roots in the type of regionalism developed by the sociologist Howard W. Odum and in regional sociology, human ecology, and urban sociology. The study of social structure is common to both disciplines, but the regional scientist is more insistent that such study be intimately linked to the analysis of economic structure and the pattern of cities and central places over space, and in a way that can be systematically formulated (Whitney 1957).

Concern with transportation systems, water resource use, agricultural development, and other problems has led to joint interests and studies with transportation engineers, hydrologists, agricultural economists, and other specialists.

It is frequently stated that regional science overlaps most with geography. Both are concerned with patterns of distribution, distances, and flows over space. However, while recognizing the need for comprehensive descriptive systems, for cartographic techniques and studies, and for the need to study physical phenomena together with the social phenomena which occur in space, regional scientists have also stressed the development of abstract model systems. Recent developments in theoretical and quantitative geography have made significant contributions related to the work of regional scientists (see, for example, Bunge 1962).

Finally, regional science has a close relation with city and regional planning. Many of the problems studied by regional scientists represent issues central to the planner, but there remain differences. The planner is much closer to the policy maker in his quest for immediate recommendations. He is more concerned with procedural and design steps and has less concern with development of abstract models.

### Organization and profession

The major professional organization in the field is the Regional Science Association, which had a world-wide membership of about 2,600 in 1967, of whom 1,500 lived in the United States. The association, founded in 1954, has several regional sections in the United States. There are a number of national and language groups in Europe, Asia, Africa, and Latin America either in existence or in the process of formation.

The major work of the association is the organization of professional conferences, held annually in the United States and in Europe, and period-

ically in the Far East and in Latin America. Proceedings of the conferences are regularly published as volumes in the series Regional Science Association, *Papers and Proceedings*. In addition, the association cooperates in the distribution of a technical journal, *Journal of Regional Science*, and in the publication of a series of monographs.

Regional science is a subject of graduate study in a number of universities in the United States and abroad. A full-time degree program is offered at the University of Pennsylvania, and programs are currently being developed elsewhere.

WALTER ISARD AND THOMAS A. REINER

[Directly related are the entries AREA; CENTRAL PLACE; PLANNING, SOCIAL, article on REGIONAL AND URBAN PLANNING; REGION; SPATIAL ECONOMICS.]

#### BIBLIOGRAPHY

- ALONSO, WILLIAM 1960 A Theory of the Urban Land Market. Regional Science Association, *Papers and Proceedings* 6:149-157.
- ALONSO, WILLIAM 1964 Location Theory. Pages 78-106 in John Friedmann and William Alonso (editors), *Regional Development and Planning*. Cambridge, Mass.: M.I.T. Press.
- BERRY, BRIAN J. L.; and GARRISON, W. L. 1958 Recent Developments of Central Place Theory. Regional Science Association, *Papers and Proceedings* 4:107-120.
- BERRY, BRIAN J. L.; and FRED, ALLAN 1961 *Central Place Studies: A Bibliography of Theory and Applications*. Bibliography Series, No. 1. Philadelphia: Regional Science Research Institute. → Reprinted in 1965, with additions through 1964.
- BUNGE, WILLIAM W. 1962 *Theoretical Geography*. Lund Studies in Geography, Series C: General and Mathematical Geography, No. 1. Lund (Sweden): Gleerup.
- CARROTHERS, GERALD A. P. 1956 An Historical Review of the Gravity and Potential Concepts of Human Interaction. *Journal of the American Institute of Planners* 22:94-102.
- DACEY, MICHAEL F. 1960 Selection of an Initial Solution for the Traveling-salesman Problem. *Operations Research* 8:133-134.
- DACEY, MICHAEL F. 1963 The Status of Pattern Analysis: Identification of Problems in the Statistical Analysis of Spatial Arrangement. Regional Science Research Institute, Discussion Paper Series, No. 3. Unpublished manuscript.
- DACEY, MICHAEL F.; and TUNG, TSE-HSIUNG 1962 The Identification of Randomness in Point Patterns. *Journal of Regional Science* 4:83-96.
- DAVIDOFF, PAUL; and REINER, THOMAS A. 1962 A Choice Theory of Planning. *Journal of the American Institute of Planners* 28:103-115.
- DEUTSCH, KARL W.; and ISARD, WALTER 1961 A Note on a Generalized Concept of Effective Distance. *Behavioral Science* 6:306-311.
- FISHER, JOSEPH L. 1957 Potential Contributions of Regional Science to the Field of Economics. Regional Science Association, *Papers and Proceedings* 3:17-23.
- FRIEDMANN, JOHN R. P. 1956 The Concept of a Planning Region. *Land Economics* 32:1-13.
- FUCHS, VICTOR R. 1962 Statistical Explanations of the Relative Shift of Manufacturing Among Regions of the United States. Regional Science Association, *Papers and Proceedings* 8:105-126.
- GARNSEY, MORRIS E. 1956 The Dimensions of Regional Science. Regional Science Association, *Papers and Proceedings* 2:27-39.
- HARRIS, BRITTON 1959 Urbanization Policy in India. Regional Science Association, *Papers and Proceedings* 5:181-203.
- HEADY, EARL O.; and EGBERT, ALVIN C. 1962 Programming Models of Interdependence Among Agricultural Sectors and Spatial Allocation of Crop Production. *Journal of Regional Science* 4, no. 2:1-20.
- HENDERSON, JAMES M. 1957 The Utilization of Agricultural Land: A Regional Approach. Regional Science Association, *Papers and Proceedings* 3:99-114.
- HENDERSON, JAMES M. 1958 *The Efficiency of the Coal Industry: An Application of Linear Programming*. Cambridge, Mass.: Harvard Univ. Press.
- HERBERT, JOHN D.; and STEVENS, BENJAMIN H. 1960 A Model for the Distribution of Residential Activity in Urban Areas. *Journal of Regional Science* 2, no. 2:21-36.
- HOOVER, EDGAR M. 1948 *Location of Economic Activity*. New York: McGraw-Hill.
- HOOVER, EDGAR M.; and VERNON, RAYMOND 1959 *Anatomy of a Metropolis: The Changing Distribution of People and Jobs Within the New York Metropolitan Region*. Cambridge, Mass.: Harvard Univ. Press. → A paperback edition was published in 1962 by Doubleday.
- ISARD, WALTER 1956a *Location and Space-economy: A General Theory Relating to Industrial Location, Market Areas, Trade and Urban Structure*. Cambridge, Mass.: M.I.T. Press; New York: Wiley.
- ISARD, WALTER 1956b Regional Science, the Concept of Region, and Regional Structure. Regional Science Association, *Papers and Proceedings* 2:13-26.
- ISARD, WALTER 1957 Potential Contribution of Regional Science to the Field of Political Science. Regional Science Association, *Papers and Proceedings* 3:29-32.
- ISARD, WALTER; and ISARD, PETER 1965 General Social, Political and Economic Equilibrium for a System of Regions. Parts 1 and 2. Regional Science Association, *Papers and Proceedings* 14:1-33; 15:7-25.
- ISARD, WALTER; and REINER, THOMAS A. 1962 Regional Science and Planning. Regional Science Association, *Papers and Proceedings* 8:1-35.
- ISARD, WALTER; SCHOOLER, EUGENE W.; and VIETORISZ, THOMAS 1959 *Industrial Complex Analysis and Regional Development: A Case Study . . .* Cambridge, Mass.: M.I.T. Press.
- ISARD, WALTER; and SMITH, T. E. 1966a On the Resolution of Conflict Among Regions of a System. Regional Science Association, *Papers and Proceedings* 17:19-46.
- ISARD, WALTER; and SMITH, T. E. 1966b A Practical Application of Game Theoretical Approaches to Arms Reduction (and to Goal Determinations Among Regional Planning Authorities). Peace Research Society, *Papers* 14:85-98.
- ISARD, WALTER; and SMITH, T. E. 1967 Location Games: With Application to Classic Location Problems. Unpublished manuscript. → To appear in Volume 19 of Regional Science Association, *Papers and Proceedings*.
- ISARD, WALTER; and TUNG, TSE-HSIUNG 1964 Selected Non-economic Commodities: Definitions and Specu-



- lations on Supply and Demand, Measurement, and Utility. Regional Science Association, *Papers and Proceedings* 13:71-91.
- ISARD, WALTER et al. 1960 *Methods of Regional Analysis: An Introduction to Regional Science*. New York: Wiley; Cambridge, Mass.: M.I.T. Press.
- McLAUGHLIN, GLEN E.; and ROBOCK, STEFAN 1949 *Why Industry Moves South: A Study of Factors Influencing the Recent Location of Manufacturing Plants in the South*. Washington: National Planning Association, Committee of the South.
- MANNE, ALAN S. 1958 A Linear Programming Model of the U.S. Petroleum Refining Industry. *Econometrica* 26:67-106.
- MARBLE, D. F.; and NYSTUEN, JOHN D. 1963 An Approach to the Direct Measurement of Community Mean Information Fields. Regional Science Association, *Papers and Proceedings* 11:99-111.
- MARSCHAK, THOMAS A. 1958 *A Spatial Model of U.S. Petroleum Refining*. RAND Research Memorandum No. RM-2205. Santa Monica, Calif.: RAND Corp.
- MEYER, JOHN R. 1963 Regional Economics: A Survey. *American Economic Review* 53:19-54.
- MILLER, RONALD E. 1963 *Domestic Airline Efficiency: An Application of Linear Programming*. Cambridge, Mass.: M.I.T. Press.
- MITCHELL, ROBERT B. 1961 On Paradigms and Paradoxes: A City Planner Looks at Regional Science. Regional Science Association, *Papers and Proceedings* 7:7-15.
- NEFT, DAVID S. 1965 *Statistical Analysis for Areal Distributions*. Philadelphia: Regional Science Research Institute.
- NYSTUEN, JOHN D.; and DACEY, MICHAEL F. 1961 A Graph Theory Interpretation of Nodal Regions. Regional Science Association, *Papers and Proceedings* 7:29-42.
- OLSSON, GUNNAR 1965 *Distance and Human Interaction: A Review and Bibliography*. Regional Science Research Institute. Bibliographic Series, No. 2. Philadelphia: The Institute.
- PERLOFF, HARVEY S. et al. 1960 *Regions, Resources and Economic Growth*. Baltimore: Johns Hopkins Press.
- STEVENS, BENJAMIN H. 1961 Linear Programming and Location Rent. *Journal of Regional Science* 3, no. 2:15-26.
- STOUFFER, S. A. 1960 Intervening Opportunities and Competing Migrants. *Journal of Regional Science* 2, no. 1:1-20.
- TEITZ, MICHAEL B. 1962 Regional Theory and Regional Models. Regional Science Association, *Papers and Proceedings* 9:35-50.
- TEITZ, MICHAEL B. 1964 Regional Theory and Regional Models. Ph.D. dissertation, Univ. of Pennsylvania.
- TOBLER, WALDO R. 1963 Geographical Area and Map Projections. *Geographical Review* 53:59-78.
- VIETORISZ, THOMAS 1956 Regional Programming Models and the Case Study of a Refinery-Petrochemical-Synthetic Fiber Industrial Complex for Puerto Rico. Ph.D. dissertation, M.I.T.
- WARNTZ, WILLIAM 1959 *Toward a Geography of Price: A Study in Geo-econometrics*. Philadelphia: Univ. of Pennsylvania Press.
- WARNTZ, WILLIAM; and NEFT, DAVID S. 1960 Contributions to a Statistical Methodology for Areal Distributions. *Journal of Regional Science* 2, no. 1:47-66.
- WEBBER, MELVIN M. 1964 Culture, Territory, and the Elastic Mile. Regional Science Association, *Papers and Proceedings* 13:59-69.

- WHITNEY, VINCENT H. 1957 Potential Contributions of Regional Science to the Field of Sociology. Regional Science Association, *Papers and Proceedings* 3:24-28.
- WHITTLESEY, DERWENT S. 1954 The Regional Concept and the Regional Method. Pages 19-68 in Preston E. James and Clarence F. Jones (editors), *American Geography: Inventory and Prospects*. Syracuse Univ. Press.
- WINGO, LOWDON JR. 1961 *Transportation and Urban Land*. Washington: Resources for the Future.
- ZIPE, GEORGE K. 1949 *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology*. Cambridge, Mass.: Addison-Wesley.

## REGRESSION

See under LINEAR HYPOTHESES; see also the biography of GALTON.

## REGULATION OF INDUSTRY

In its broadest sense the term "regulation" may be taken to comprehend all of the controls that government imposes on business of all kinds. In the narrower sense in which the term is used here, it is limited to control of the services provided and rates charged by private enterprises engaged in the provision of transport, communication, electricity, gas, and other utility services. In this sense, regulation is largely an American institution. Elsewhere in the world, industries rendering these services are usually owned and operated by government. Experience with regulation of such industries is thus confined, in the main, to the United States.

At one time, economists sought to distinguish industries that should be regulated by pointing out characteristics that made them naturally monopolistic—overcapacity, high fixed costs, a tendency toward destructive competition and eventual collocation—and attributed regulation to the consumer's disadvantage in buying from monopolists. But today there are industries possessing these characteristics that are not regulated, as well as other monopolies that are not controlled, while transport, though increasingly competitive, has been increasingly controlled. At one time, too, the courts excepted the transport and utility industries from their general disapproval of regulation on the ground that these industries were peculiarly "affected with a public interest." But this distinction was abandoned as long ago as 1934. No hard and fast line can now be drawn, in economics or in law, between the industries that should be regulated and those that should not.

The regulatory commissions. Regulation was first attempted through provisions written into statutes, ordinances, and franchises permitting the use of city streets. But this method of control proved

to be clumsy and ineffective. Legislatures then turned to the creation of administrative agencies, called commissions, charging them with responsibility for the enforcement of controls. The first of these bodies were set up by the states in the nineteenth century to regulate the railroads. The first such federal agency, the Interstate Commerce Commission (ICC), created in 1887, was also empowered to regulate this industry, and in 1914, when the courts upheld the right of the ICC to control intrastate as well as interstate operations, that commission came to dominate the regulation of railroads, and the state commissions were relegated to a minor role. At the same time, with the growing importance of electricity, telephone, and other utility services, the state commissions were given jurisdiction over these fields, being transformed from railroad commissions into bodies known as public utility or public service commissions. Such an agency is found in nearly every state today.

During the 1930s the federal government set up three other commissions with power to regulate transport and utility services and rates: the Federal Power Commission (FPC) with jurisdiction over hydroelectric installations on navigable waterways and over the interstate transmission of electricity and natural gas, the Federal Communications Commission (FCC) with jurisdiction over interstate and international telephone and telegraph services, and the Civil Aeronautics Board (CAB) with jurisdiction over civil aviation. Some of the duties of these agencies do not conform to the usual pattern of regulation. The FPC is empowered not only to fix the rates charged for the transmission of gas through interstate pipelines but also to regulate the price charged for gas at the wellhead. The FCC is required to allocate channels for broadcasting and to exercise some control over the character of broadcasts but does not regulate the prices charged to advertisers by the broadcasters. There are other federal commissions whose functions do not fall within the field of regulation, as narrowly defined. The Securities and Exchange Commission, for instance, and the Federal Trade Commission are concerned with the prevention of fraud but not with the control of services and rates.

The regulatory commission differs in form and in character from other agencies of government. The typical commission is composed of an odd number of members appointed by the executive, with representation assured to each of the major political parties, with terms long and overlapping, tenure legally protected, and decisions subject not to executive veto but only to judicial review. Its specialization and its independence are designed to

put responsibility for regulation in the hands of experts, to permit the use of informal procedures, and thus to speed decisions. Its multiple membership and its overlapping terms are intended to assure soundness of judgment and continuity of policy. Its bipartisanship and its protection from political pressures are expected to ensure impartiality. The commission has sometimes been characterized as a fourth branch of government, not clearly to be classified as legislative, executive, or judicial. Exercising a function delegated to it by the legislature, it gives detailed substance to general policies prescribed by law. Located in the executive establishment, it is subject to administrative directives and constraints. Deciding disputed cases on the basis of general principles, it acts, in essence, as a court.

**The task of regulation.** It is the central purpose of regulation to prevent the owners of regulated enterprises from obtaining monopoly profits. To this end the commissions control the rates that may be charged. But there are other activities that must also be controlled: entry into the industry, the availability, quality, and safety of service, financial practices, and accounting methods.

Entry of competing enterprises is restricted in order to safeguard the financial strength of those already in the field. Control over service is necessitated by the absence of competition. Extension of service may be required and abandonment of particular services forbidden. Standards to govern the quality of service may be established, though initiative here is usually left to industry, with the commissions acting only in response to customer complaints. Safety requirements, of minor importance in some fields, play a major role in the regulation of such industries as civil aviation and the generation of electricity from nuclear energy.

Control of financial operations is undertaken to maintain the credit of regulated companies and thus to assure their continued ability to attract capital and render service. In most cases, securities may not be issued without commission approval. Overcapitalization can thus be forestalled and a balanced capital structure required. Padding of costs and diversion of earnings are prevented by supervising transactions between related concerns.

Control of accounting is essential to the administration of the other controls. Rates cannot be effectively regulated, proposals for extension or abandonment properly appraised, or financial operations adequately supervised in the absence of accurate knowledge regarding company accounts. Commissions are therefore generally empowered to prescribe methods of accounting and to require periodic reports.



The regulation of rates has two aspects: control of the rate level and control of the rate structure. The rate level has to do with the size of a company's earnings. Here, the commission must seek to attain equity as between investors and consumers. The rate structure has to do with the particular rates paid by different classes of customers for different units of service. Here, the commission must seek to prevent treatment that is unduly discriminatory, favoring some customers and harming others.

**The rate level.** The general level of rates should be high enough, first, to cover the current costs of operation and, second, to yield an adequate return on the investment. Ideally, this return should be sufficient to be fair to past investors, avoiding impairment of the value of their property. It should also be high enough to preserve the credit standing of the regulated company so that it will be able to attract the capital needed if its service is to be improved and expanded.

Control of the rate level involves three different processes. First, it is necessary to control operating expenses. These expenses are usually the largest sum that must be covered in the rates. If they are inflated, the rates will be too high. Second, it is necessary to determine the rate of return on investment that is to be taken as fair to investors and as adequate for the maintenance of credit. This rate is expressed as a percentage of the value of the company's property. Third, it is necessary to determine the value of this property, estimating construction or reproduction costs and making due allowance for depreciation.

The commissions have not been vigorous in controlling operating expenses. In general, they have taken the view that decisions respecting the wisdom of expenditures lie within the province of management and have refused to substitute the judgment of regulators for that of managers. Accordingly, they have never sought the power to require that budgets be submitted for approval before money could be spent. The most that they do is to examine reports on expenses incurred in the past, and, if certain items seem improper, they may warn the companies concerned that they will not include such items in computing future rates. But regulation may be so lax that such warnings will have little effect.

**The rate base.** In determining the rate of return to be allowed on investments, the commissions operate within the limits of the constitution as it is interpreted by the courts. Under the fifth and fourteenth amendments to the constitution, federal and state governments, respectively, are

forbidden to deprive any person of property without due process of law. Applying this prohibition, the courts undertook, for many years, to review the rates that the commissions set. Their concern was to protect investors; their attention was focused, accordingly, upon the adequacy of earnings. Investors in regulated industries, they held, should be permitted a fair return upon a fair value of their investments. But the issues usually brought before them related not to the percentage adopted as a fair return but to the method used in determining fair value. Valuation of the rate base thus came to be the heart of the problem of controlling rates.

The points at issue in the valuation controversy had to do, in part, with the propriety of including intangible assets in the rate base and with the methods used in the treatment of depreciation. But these matters were overshadowed by the dispute between those who held that the properties should be evaluated on the basis of their original cost and those who would evaluate them on the basis of reproduction cost. With prices rising, reproduction cost would necessitate a higher level of earnings and of rates.

Beginning in 1898 the courts insisted that some weight be given to reproduction cost, rejecting a number of rate determinations on the ground that the weight accorded was insufficient. But they never stated what the proper weight would be, accepting a lower weight as adequate in one case and rejecting a higher weight as inadequate in another. As a result, rate determinations were held up by protracted litigation. The commissions, instead of taking formal action, sought to obtain such rate concessions as they could through informal negotiation. For nearly half a century the effectiveness of regulation was seriously impaired.

This situation was ended in 1944, when the Supreme Court declared that the method of valuation was to be left to the discretion of the commissions and that consideration, accordingly, need no longer be given to reproduction cost. Since that time, valuations made by the federal commissions and most of the state commissions have been based on original cost. Fair value, however, is still based on reproduction cost in a minority of states.

**The rate of return.** With settlement of the controversy over valuation, more importance has come to be attached to determination of a fair rate of return. Here, recognition has been given, first, to the fact that earnings need to be high enough to enable regulated companies to attract new capital. But no rules have been established for ascertaining the level of earnings required for this purpose. The nature of a company's capital structure, though

relevant, is not taken into account. The return required to ensure a market for common stocks is uncertain. Earnings can always be set at a level that will make such stocks attractive, but this may involve a larger return than is in fact required. Recognition is given, second, to the need for a return that will be fair to past investors. Here, the courts have said that earnings will be fair if they are equal to those obtainable in other industries with comparable risks. But this rule is rendered inoperable by the absence of industries with similar risks. The earnings needed to attract new capital, moreover, are likely to differ from those required for fairness to past investors. According to the one standard or the other, the return allowed will thus be too high or too low.

Other issues raised by the rate of return have been ignored. Earnings might well be averaged over time; they are calculated for a single year. Earnings might be adjusted to afford an incentive to efficiency; such adjustments have been made, but they are exceptions to the general rule. In practice, the commissions and the courts announce the percentages of return they hold to be fair, running from less than 6 per cent to as much as 8 per cent (or more in the case of airlines)—these figures varying from state to state, industry to industry, and year to year. But they do not disclose the considerations that have led to their decisions. The principles, if any, that govern the determination of a fair return are thus unknown.

*The level of transport rates.* In controlling utility earnings, fair return on fair value is still the rule, but, in the case of transport, this rule has long since been abandoned. Public utilities, in general, enjoy monopolies of essential services. Utility rates designed to yield a specified return can usually be relied upon to do so. Railroads, on the other hand, have always competed with one another. A fair return could not be separately fixed for each competitor but had to be established as an average for those in a competing group. Railroads have come to be faced, moreover, with the competition of other transport media. There is no assurance that a return adopted as fair for them will actually be obtained.

In 1920 Congress directed the ICC to set railway rates at a level that would enable the roads to earn an average return of  $5\frac{1}{2}$  per cent. This goal was never attained. In 1933 Congress enacted, in the Interstate Commerce Act, a new rule of rate making. The commission was now directed "to give due consideration . . . to the effect of rates on the movement of traffic; to the need . . . of adequate and efficient transportation service at

the lowest cost consistent with the furnishing of such service; and to the need of revenues sufficient to enable the carriers . . . to provide such service" (sec. 15a). The new rule made no mention of fair value. It called only for a return that would enable the railroads to attract new capital. Recognizing the existence of growing competition with increasing elasticity of demand, it acknowledged that the desired level of earnings in transport might not be attainable by raising rates. This rule was applied, within the next few years, not only to railroads but also to carriers by highway, by water, and by air.

*The rate structure.* Regulation is applied to the rate structure in order to control discrimination among consumers. The rates charged by transport and utility enterprises for various units of service are highly differentiated; railroads, for instance, charge different rates per ton-mile for carrying different commodities, or for the same commodity between different points; electric companies charge different rates for different uses and for different hours. These differences are to be explained, in part, by differences in cost. But they are also to be attributed, in large measure, to differences in demand. Where demand is elastic, rates are low; where it is inelastic, they are high.

Under certain circumstances there is a case to be made for discrimination. There must be idle capacity so that output can be increased. There must be high fixed charges so that cost per unit can be reduced as output grows. The added output must be salable only at the lower rates. These rates must be high enough to contribute to overhead. The general level of rates must be regulated to keep earnings reasonable and discrimination within bounds. Given these conditions, discrimination may result in lower prices for each class of customers and for each block of sales. It may bring about a fuller utilization of resources and a wider consumption of services.

Discrimination by regulated utility monopolies may thus have a generally beneficial effect. Discrimination by competing transport companies is unlikely to do so. The establishment of lower rates for goods that otherwise would not move at all may be of general benefit. But rates are also set lower on hauls where railroads compete for traffic and higher on hauls where they do not. Traffic that is attracted by the lower charges is diverted to one carrier from another. The one gets more revenue; the other gets less. With the former, unit costs are lowered; with the latter, they are raised. Shippers who use competing lines are benefited; those who use noncompeting lines are handi-



capped. It cannot be said that such discrimination is of benefit to the community as a whole.

*Regulation of the rate structure.* The law does not forbid discrimination as such but only that discrimination which may be held to be undue or unreasonable, giving one user of service an unfair competitive advantage and placing another under an unfair handicap. It is thus the function of the commissions to determine when discrimination is to be permitted and when forbidden. This work has taken a large part of the time of the ICC. It has been given little attention by the state utility commissions.

Initiative in setting individual rates is taken by the regulated companies, the commissions acting only in response to customer complaints. In the case of rail transport, such complaints have been numerous. In dealing with them over the years, the ICC has developed a body of precedent to which it adheres. In general the commission has sought to prevent discrimination among competing shippers. But it has also permitted discrimination where this was necessary to enable one carrier to compete with another for particular commodities and hauls. When the two principles have been in conflict, the latter has prevailed. The resulting structure of rates has been highly discriminatory.

The rate structure established by the utility companies has typically involved the use of classifications and differentials that have been based in part on differences in cost but in larger part on differences in the elasticity of demand. In the case of electricity, for instance, the low rates have gone to industry, where demand was elastic because it could produce its own power, and the high rates have been charged to householders, whose demand was inelastic because they could not. The commissions have accepted this pattern, subjected it to little scrutiny, and made little effort to have it modified.

*Regulation of utility monopolies.* In general, the regulation of utility monopolies has been lax. Control of service and the structure of rates has been minimal. Attention has been directed primarily to the rate level. But here, operating expenses have not really been controlled. The rate base, long inflated by reproduction-cost valuations, may now be tied to original cost. But the rate of return allowed is determined arbitrarily, without reference to stated criteria. Control of earnings is neither tight nor continuous. The earnings permitted have been well above those required to enable the utility companies to attract new capital. A rate case, moreover, may be brought as infre-

quently as once in a decade. In the meantime, if demand has grown and costs have declined, excessive profits have been realized and retained.

*Regulation of transport competition.* In the case of transport, regulation has come to have a different purpose and effect. Here, monopoly has given way to competition. It should have been possible, accordingly, to relax controls. Instead, they were intensified. It came to be the purpose of regulation not to protect shippers against monopolistic carriers but to protect competitive carriers against one another. The pattern of control developed for the railroads was extended to the other modes of transport. Control of entry and of rates came to be employed to make sure that each of these modes retained its accustomed share of the traffic. Entry into the trucking business was curtailed and routes were restricted. Emphasis in the control of rail rates shifted from the imposition of maxima to the imposition of minima. The railroads, in practicing discrimination, were permitted to meet the lower rates of other carriers but not to undercut them, even though still lower figures would have more than covered their out-of-pocket costs. As a result, the allocation of traffic came to be influenced less by the comparative economic advantages of the different transport media than by the regulations imposed by the ICC. It was not until 1958 that policy began to move in the direction of relaxing controls so that competition could be given somewhat freer play.

*Regulation of civil aviation.* Control of civil aviation is a case by itself. Here, safety regulation is in the hands of the Federal Aviation Agency and economic regulation in the hands of the Civil Aeronautics Board (CAB). It is with the latter that we are here concerned.

Air carriers, unlike surface carriers, derive their revenues predominantly from passengers. Until after World War II, airline revenues fell short of expenses and all the lines were subsidized. The government was concerned less with the regulation of fares than with determination of the subsidies. After the war, however, the trunk (long-haul passenger) lines gradually went off subsidy and the CAB undertook to develop a policy to govern regulation of the level of fares. Under this policy, as made public in 1960, the trunk lines are to be allowed a weighted average return of 10.5 per cent on their investment; the local feeder lines from 9 to 12.5 per cent depending upon the character of their capital structures, with 5.5 per cent allowed on their bonds and 21.35 per cent on their common stock. It is obvious that such rates of

return will not make for tight control of earnings. As for the structure of fares, no study has yet been made and no policy announced.

The CAB has maintained strict control over entry, permitting no new trunk lines to be established since it was created in 1938. It has encouraged the creation of local feeder lines, keeping them in existence by subsidization. Most of the board's time has been devoted to passing on applications for the right to fly particular routes. Here it has sought to ensure the existence of competitive services and, by granting added routes to weaker companies, to equalize the profitability of the airlines. The board has used its control of routes to stimulate the modernization of equipment and increase the frequency and speed of flights. At the same time, its control has delayed adaptation of routes to changing needs and has discouraged diversity in types of service. It may be questioned whether the resulting pattern of routes and complement of services is superior to those that would have obtained in the absence of regulation.

*Regulation of broadcasting.* In the case of radio and television, the Federal Communications Commission (FCC) is concerned primarily with the allocation of channels for broadcasting. Licenses conveying rights to exclusive use of channels, their value running into millions of dollars, are granted to broadcasters without charge. These grants are made on the basis of stated criteria. But the criteria are imprecise, subjective, inconsistent, and inoperable. They are given lip service but in practice are ignored. The real grounds for action are not disclosed. Commissioners have been subjected to improper pressures and have succumbed, in some cases, to bribery. Critics of the licensing process have proposed that licenses, instead of being given away, be auctioned off to the highest bidder. Such a change would put an end to the ambiguity and the temptation to corruption inherent in the present method of licensing. But, in itself, it would do nothing to improve the quality of programming.

It is here that regulation has failed most seriously. Its licensing power gives the FCC control over the structure of the industry, over contractual relations between the networks and their affiliated stations, and over progress in the adoption of new technology, but only nominal control over the character of programming. The commission has reserved a number of channels for educational broadcasting, authorized experimentation with pay television, and sought to curb the broadcasting of lotteries and of obscenity. But it has not revoked licenses or refused renewals on the ground that

the programs presented have failed to fulfill the broadcaster's promise to serve the public interest. Commission chairmen have admonished the industry from time to time to raise the general quality of their programs and to avoid the crasser abuses of advertising. But in the face of powerful political pressure, the commission has lacked the courage to order them to do so.

*Shortcomings of regulation.* There are a few regulatory commissions that have consistently maintained a record of vigorous action in the public interest and a few that have taken such action upon occasion. But, by and large, the agencies have fallen short of the performance they had promised. They have not anticipated changing conditions or developed broad policies to guide control but have waited for cases to be brought before them, permitting policy to emerge from the processes of adjudication. They have not sped decisions through informal action but have delayed them by adhering to cumbersome procedures. They have not fulfilled their promise of independence and impartiality. The commissions are dependent on the executive for appointments, for budget requests, and for political support; on the legislature for jurisdiction, for powers, for confirmation of appointments, and for appropriations. Lacking insulation from political pressures, they have been unable to take an independent line.

More than this, the commissions have come increasingly to serve the interests of the industries they were set up to regulate. When first established, through popular demand for reform, they had the support of the executive and the legislature and thus were able to embark with some zeal upon their regulatory duties. In time, however, public attention has shifted to other problems. Executives have made poor appointments and have failed to provide political backing. Legislatures have denied adequate jurisdiction, powers, and appropriations. The commissions, lacking any other clientele, have turned to the regulated industries for support. Working day in and day out with industry officials, they have become immersed in industry's problems and have come to share its point of view. The consumer's interest has been lost from sight.

*Inherent limitations of regulation.* The commissions' powers could be strengthened, their personnel improved, and their appropriations increased. But there are limitations inherent in the pattern of regulation that such reforms could not remove. Regulation cannot itself prescribe quality, force efficiency, or require innovation; to do so, it would have to duplicate the central functions of



management. Nor does it offer a reward for good performance or impose a penalty for poor performance in these respects. On the contrary, rate regulation involves cost-plus pricing. If management is diligent in cutting costs, rates will be reduced; if it is indolent, they will be increased. If regulation were tight there would be no incentive to efficiency or to progress. If such an incentive does, in fact, exist, it is because regulation is lax.

Regulation could conceivably prevent monopolistic enterprises from charging monopoly prices and obtaining monopoly profits. But it does not attempt to maximize consumption by setting rates at the lowest level that will yield a fair return. Regulation is backward-looking. The necessary level of rates is computed by multiplying price per unit by past demand. Elasticity of demand is not taken into account. The possibility that lower charges might be justified by growth of demand or by innovations in technology is ignored. As a result, the level of rates, even if monopoly profits are excluded, is higher than it needs to be.

Regulation is inevitably slow. It must satisfy the requirements of due process: investigate, give notice, hold hearings, study the record, make findings, issue orders, permit appeals. All this takes time and delays action. In some cases, delay may be harmful, as when it permits earnings to rise well above or to fall far below the return required to attract new capital. In other cases, it may be helpful, as when it brakes an inflationary spiral of wages and prices. But here, the merit of regulation lies not in its efficiency but in its inefficiency.

Regulatory agencies tend to become industry-minded, seeking to protect the regulated industry against low rates and impaired earnings. This orientation makes, too, for regulation-mindedness. Where effectiveness of the regulator's controls is weakened by the freedom of firms beyond his jurisdiction, he seeks wider jurisdiction. Where the applicability of his controls is lessened by changing circumstances, he seeks tighter controls. Thus, in transport, where monopoly has given way to competition, the original purpose of regulation might have been better served if jurisdiction had been narrowed or controls relaxed. But policy, over the years, has favored expansion of coverage and elaboration of detail.

The regulated industry comes, in the end, to have two masters: its own management and the regulatory agency. Managerial decisions are reviewed. Where the regulatory agency finds them to be wise, it allows them to stand. Where it finds them to be unwise, it exercises a veto power. Negotiation between the managers and the regula-

tors takes time; decisions are inevitably delayed. Division of authority diverts energy from the solution of external problems to the prosecution of internal disputes. At the same time, it dissipates responsibility.

The ICC is required, for instance, to consider the probable effect of the rates it sets upon the volume of traffic. In doing so, it may substitute its own judgment for that of railway managements. If earnings fall, there is no one whom the owners of the roads can hold responsible. The managements are deprived of power; the commission has power but lacks responsibility. Even if managements are not reversed, decisions may be so delayed that substantial losses are sustained. In cases such as this, the management escapes responsibility by blaming the regulators. But the regulators cannot be held accountable.

CLAIR WILCOX

[Directly related are the entries COMMISSIONS, GOVERNMENT; PRICES, article on PRICING POLICIES, TRANSPORTATION, article on ECONOMIC ASPECTS. Other relevant material may be found in WELFARE STATE.]

#### BIBLIOGRAPHY

- BERNSTEIN, MARVER H. 1955 *Regulating Business by Independent Commission*. Princeton Univ. Press.  
 BONBRIGHT, JAMES C. 1961 *Principles of Public Utility Rates*. New York: Columbia Univ. Press.  
 CARY, WILLIAM L. 1967 *Politics and the Regulatory Agencies*. New York: McGraw-Hill.  
 CAVES, RICHARD E. 1962 *Air Transport and Its Regulators: An Industry Study*. Harvard Economic Studies, Vol. 120. Cambridge, Mass.: Harvard Univ. Press.  
 LOCKLIN, DAVID P. (1935) 1966 *Economics of Transportation*. 6th ed. Homewood, Ill.: Irwin.  
 PHILLIPS, CHARLES F. JR. 1965 *The Economics of Regulation*. Homewood, Ill.: Irwin.

#### REHABILITATION

For material on physical rehabilitation, see BLINDNESS; HEARING; LANGUAGE, article on SPEECH PATHOLOGY. Mental rehabilitation is described in the articles under MENTAL DISORDERS, TREATMENT OF; PSYCHIATRY; PSYCHOANALYSIS. For neighborhood rehabilitation, see PLANNING, SOCIAL, article on REGIONAL AND URBAN PLANNING. Other aspects of rehabilitation are discussed in HOUSING; PENOLOGY, article on PROBATION AND PAROLE; SOCIAL WORK; VOCATIONAL REHABILITATION.

#### REICH, WILHELM

Wilhelm Reich was born in 1897, in what was the Austrian part of Poland, and died in 1957 in Lewisburg, Pennsylvania. He was a psychoanalyst and social philosopher, originating one of the dissi-

dent trends of thought derived from Freudian psychoanalysis. His early contributions to the clinical field are accorded general esteem, while his later ideas (which encompass virtually all the sciences from physics to psychology), his social philosophy, and his attempts to translate this philosophy into large-scale social action are generally considered grandiose, the products of a passion that gradually gave way to schizophrenic deterioration.

Reich became a dissident from Freudian orthodoxy by following Freud's original path to its logical end, as he saw it. His productive thought took off from two of Freud's early hypotheses. The first maintains the existence of a type of neurosis (*Aktualneurose*) caused by an actual physiological disturbance of sexuality. According to this view, frustrated excitement, coitus interruptus, etc., can independently of any mental factors produce this kind of neurosis. Although Freud himself continued to consider the hypothesis correct, he did not pursue this trend of thought once he became involved in the investigation of psychological conflicts. The second hypothesis maintains that the characteristics of a person's sexuality determine the characteristics of his personality. This is an oversimplification that Freud implicitly withdrew when he hypothesized the ego as an entity to be studied in its own right.

Reich extended Freud's hypothesis of the *Aktualneurose*. He held that not only the *Aktualneurose* but also the psychoneuroses are characterized, indeed caused, by a damming up of undischarged and, in this state, noxious sexual energy. The dissipation of noxious sexual energy is the irreplaceable function of the orgasm. Extending Freud's hypothesis that sexuality determines personality, Reich declared that psychic health depends upon orgasmic potency. The orgasmically potent individual is free of destructive aggression and will spontaneously, without the pressure of a moral conscience, enjoy doing what is right and socially beneficial (1927, vol. 1).

The goal of psychoanalytic therapy for Reich, therefore, is to free the orgasmic function of the neurotic. Reich became convinced that in order to achieve this goal, psychoanalytic therapists must above all attack the patient's resistances rather than interpret unconscious contents, as they typically did at that time. Taking a productively fresh and wide view of character resistances, he perceived them not only in symptoms, inhibitions and anxieties, attitudes and values, but also in such less obvious guises as habits, mannerisms, and particularly muscular tensions. These last he came to regard as materialization of the "character armor,"

which function to prevent the outbreak of neurotic anxiety caused by the undischarged sexual tensions (1933-1935).

If the neuroses are maintained by the "muscular armor," then a direct attack may be made on them by melting down the "armor." This can be done by encouraging a patient to relax strategic groups of muscles. The analyst then has to assist the patient in the assimilation and mastery of emotions presumably thus liberated. Reich designated such treatment "vegetotherapy" and for a while thought that it provided him with a simple technique for the sexual liberation of individuals (1933-1935).

But by the late 1920s his interest had already gone beyond therapy merely for individuals. He sought—and believed he had found—a theory that justified large-scale social action. As a radical interpreter of the early Freud he had no doubt that suppression of the sexual impulses in children and adolescents is responsible for the high incidence of neurosis. This suppression he saw as bound up with the institution of the authoritarian family, on which the entire authoritarian structure of society in turn depended. Reich hoped he could destroy authoritarianism both in the family and in society at large by encouraging the young victims of the system to revolt and to win sexual freedom (1932). He attempted in vain to gain acceptance for these views, first in the socialist, then in the communist parties of Europe, and finally founded his own organization to propagate them.

Eventually Reich's interests took a different direction. He believed he had discovered the physical reality corresponding to "psychic energy." He called it "orgone" (1927) and claimed that the surface of the healthy human body radiated this form of energy, whereas bodies of the neurotics and those suffering from a great variety of other illnesses, including cancer, were deficient in it. In the United States, where he lived from 1939 until his death, he constructed and sold "orgone boxes" and recommended them for cancer therapy. Patients seated in them presumably absorbed life-enhancing rays, which the boxes gathered from the surrounding abundance of orgone in the universe. An agency of the U.S. government, ignorant of Reich's background and assuming his orgone to be simple commercial quackery, eventually caused him to be sent to prison, where he died of a heart attack.

During his years in America, Reich was followed and supported by devoted believers, including physicians, laymen, and artists. At the time of this writing, 1965, the Reich cult seems to be in decline but not extinguished.

PAUL BERGMAN



[For the historical context of Reich's work, see **PSYCHO-ANALYSIS**, article on **CLASSICAL THEORY**; and the biography of **FREUD**. For discussion of the subsequent development of Reich's ideas, see **SEXUAL BEHAVIOR**, articles on **SEXUAL DEVIATION**.]

#### WORKS BY REICH

- (1927) 1948 *The Discovery of the Orgone*. 2d ed., 2 vols. New York: Orgone Institute Press. → First published in German. Volume 1: *The Function of the Orgasm*. Volume 2: *The Cancer Biopathy*.  
 (1930) 1945 *The Sexual Revolution*. 3d ed. New York: Orgone Institute Press. → First published as *Die Sexualität im Kulturkampf*.  
 1932 *Der sexuelle Kampf der Jugend*. Copenhagen: Verlag für Sexualpolitik.  
 (1933-1935) 1961 *Character-analysis*. 3d ed., enl. New York: Noonday Press. → First published in German. Contains the translations of *Charakteranalyse* and *Psychischer Kontakt und vegetative Strömung*.

### REINFORCEMENT

See under **LEARNING**.

### RELATIVISM, CULTURAL

See under **CULTURE**.

### RELATIVISM, ETHICAL

See **ETHICS**, article on **ETHICAL SYSTEMS AND SOCIAL STRUCTURES**.

### RELIABILITY

See **ERRORS**; **EXPERIMENTAL DESIGN**; **PSYCHOMETRICS**; **QUALITY CONTROL**, **STATISTICAL**.

## RELIGION

The articles under this heading describe three approaches to the topic of religion in the social sciences. The history and doctrines of the major religions are discussed in **BUDDHISM**; **CHRISTIANITY**; **HINDUISM**; **ISLAM**; **JUDAISM**. The organizational aspects of religion are described in **MONASTICISM**; **PSYCHIATRY**, article on the **RELIGIO-PSYCHIATRIC MOVEMENT**; **RELIGIOUS ORGANIZATION**; **RELIGIOUS SPECIALISTS**; **SECTS AND CULTS**. Various aspects of religious belief and practice are reviewed in **CANON LAW**; **CIVIL DISOBEDIENCE**; **DEATH**; **MILLENARIISM**; **MORAL DEVELOPMENT**; **MYTH AND SYMBOL**; **NATIVISM AND REVIVALISM**; **POLLUTION**; **RELIGIOUS OBSERVANCE**; **RITUAL**. Other relevant material may be found in **CHARISMA**; **VOLUNTARY ASSOCIATIONS**. Biographical articles that are of relevance to the social scientific study of religion include **AQUINAS**; **AUGUSTINE**; **BECKER**; **BLOCH**; **BUBER**; **CALVIN**; **COMTE**; **DURKHEIM**; **ERASMUS**; **FEVRE**; **FRAZER**; **FREUD**; **HALÉVY**; **HUIZINGA**;

**JAMES**; **JUNG**; **LOWIE**; **LUTHER**; **MALINOWSKI**; **MARETT**; **MARSILIUS**; **MARX**; **MAUSS**; **MONTESQUIEU**; **MOONEY**; **OCKHAM**; **RANK**; **SELIGMAN**, C. G.; **SMITH**, **WILLIAM ROBERTSON**; **SPINOZA**; **TAWNEY**; **TROELTSCH**; **TYLOR**; **WEBER**, **MAX**.

- I. ANTHROPOLOGICAL STUDY
- II. THE SOCIOLOGY OF RELIGION
- III. PSYCHOLOGICAL STUDY

Clifford Geertz  
Robert N. Bellah  
James E. Dittes

#### I

#### ANTHROPOLOGICAL STUDY

The anthropological study of religion has been highly sensitive to changes in the general intellectual and moral climate of the day; at the same time, it has been a powerful factor in the creation of that climate. Since the early discussion by Edward Tylor, interest in the beliefs and rituals of distant, ancient, or simpler peoples has been shaped by an awareness of contemporary issues. The questions that anthropologists have pursued among exotic religions have arisen from the workings—or the misworkings—of modern Western society, and particularly from its restless quest for self-discovery. In turn, their findings have profoundly affected the course that quest has taken and the perspective at which it has arrived.

Perhaps the chief reason for the rather special role of comparative religious studies is that issues which, when raised within the context of Western culture, led to extreme social resistance and personal turmoil could be freely and even comfortably handled in terms of bizarre, presumably primitive, and thus—also presumably—fanciful materials from long ago or far away. The study of “primitive religions” could pass as the study of superstition, supposedly unrelated to the serious religious and moral concerns of advanced civilization, at best either a sort of vague foreshadowing of them or a grotesque parody upon them. This made it possible to approach all sorts of touchy subjects, such as polytheism, value relativism, possession, and faith healing, from a frank and detached point of view. One could ask searching questions about the historicity of myth among Polynesians; when asked in relation to Christianity, these same questions were, until quite recently, deeply threatening. One could discuss the projection of erotic wishes found in the “totemic” rites of Australian aborigines, the social roots and functions of African “ancestor worship,” or the protoscientific quality of Melanesian “magical thought,” without involving oneself in polemical debate and emotional distress. The application of the comparative method—the essence of anthropological thought—to religion per-

mitted the growth of a resolutely scientific approach to the spiritual dimensions of human life.

Through the thin disguise of comparative method the revolutionary implications of the work of such men as Tylor, Durkheim, Robertson Smith, Freud, Malinowski, and Radcliffe-Brown soon became apparent—at first mainly to philosophers, theologians, and literary figures, but eventually to the educated public in general. The meticulous descriptions of tribal curiosities such as soul loss, shamanism, circumcision, blood sacrifice, sorcery, tree burial, garden magic, symbolic cannibalism, and animal worship have been caught up in some of the grander intellectual battles of the last hundred years—from those over evolutionism and historicism in the late nineteenth century to those over positivism and existentialism today. Psychoanalysts and phenomenologists, Marxists and Kantians, racists and egalitarians, absolutists and relativists, empiricists and rationalists, believers and skeptics have all had recourse to the record—partial, inconsistent, and shot through with simple error as it is—of the spiritual life of tribal peoples to support their positions and belabor those of their opponents. If interest in “primitive religion” among savants of all sorts has been remarkably high, consensus concerning its nature and significance has not.

At least three major intellectual developments have exercised a critical influence on the anthropological study of religion: (1) the emergence, in the latter half of the nineteenth century, of history as the sovereign science of man; (2) the positivist reaction against this sovereignty in the first decades of the twentieth century and the radical split of the social sciences into resolutely psychological approaches, on the one hand, and resolutely sociological ones, on the other; and (3) the growth, in the interwar period, of a concern with the role of ideational factors in the regulation of social life. With the first of these came an emphasis on the nature of primitive reasoning and the stages of its evolution into civilized thought. With the second came an investigation of the emotional basis of religious ritual and belief and the separate examination of the role of ritual and belief in social integration. The concern with value systems and other features of the ideational realm led to an exploration of the philosophical dimensions of religious ideas, particularly the symbolic vehicles in terms of which those ideas are expressed.

### Evolutionism and its enemies

Like so much else in anthropology, the study of the religious notions of primitive peoples arose

within the context of evolutionary theory. In the nineteenth century, to think systematically about human affairs was to think historically—to seek out survivals of the most elementary forms and to trace the steps by which these forms subsequently developed. And though, in fact, Tylor, Morgan, Frazer, and the rest drew more on the synthetic social-stage theories of such men as Comte and Hegel than on the analytic random-variation and natural-selection ideas of Darwin, the grand concept of evolution was shared by both streams of thought: namely, that the complex, heterogeneous present has arisen, more or less gradually, out of a simpler, more uniform past. The relics of this past are still to be found scattered, like Galápagos turtles, in out-of-the-way places around us. Tylor, an armchair scholar, made no “voyage of the *Beagle*.” But in combing and organizing the reports of missionaries, soldiers, and explorers, he proceeded from the same general premise as did Darwin, and indeed most of the leading minds of the day. For them a comprehensive, historically-oriented comparison of all forms of a phenomenon, from the most primitive to the most advanced, was the royal road to understanding the nature of the phenomenon itself.

In Tylor's view, the elementary form out of which all else developed was spirit worship—*animism*. The minimal definition of religion was “a belief in spiritual beings.” The understanding of religion thus came down to an understanding of the basis upon which such a belief arose at its most primitive level. Tylor's theory was intellectualistic. Belief in spirits began as an uncritical but nonetheless rational effort to explain such puzzling empirical phenomena as death, dreams, and possession. The notion of a separable soul rendered these phenomena intelligible in terms of soul departure, soul wandering, and soul invasion. Tylor believed that the idea of a soul was used to explain more and more remote and hitherto inexplicable natural occurrences, until virtually every tree and rock was haunted by some sort of gossamer presence. The higher, more developed forms of “belief in spiritual beings,” first polytheism, ultimately monotheism, were founded upon this animistic basis, the ur-philosophy of all mankind, and were refined through a process of critical questioning by more advanced thinkers. For this earnest Quaker the religious history of the world was a history of progressive, even inevitable, enlightenment. [See TYLOR.]

This intellectualistic, “up from darkness” strain has run through most evolutionist thought about religion. For Frazer, a nineteenth-century figure who lived for forty years into the twentieth century



without finding it necessary to alter either his views or his methods, the mental progress involved was from magic to religion to science. Magic was the primordial form of human thought; it consisted in mistaking either spatiotemporal connection ("sympathetic magic," as when drinking the blood of an ox transfers its strength to the drinker) or phenomenal similarity ("imitative magic," as when the sound of drumming induces thunderheads to form) for true scientific causality. For Durkheim, evolutionary advance consisted in the emergence of specific, analytic, *profane* ideas about "cause" or "category" or "relationship" from diffuse, global, *sacred* images. These "collective representations," as he called them, of the social order and its moral force included such sacra as "mana," "totem," and "god." For Max Weber, the process was one of "rationalization": the progressive organization of religious concern into certain more precisely defined, more specifically focused, and more systematically conceived cultural forms. The level of sophistication of such theories (and, hence, their present relevance) varies very widely. But, like Tylor's, they all conceive of the evolution of religion as a process of cultural differentiation: the diffuse, all-embracing, but rather unsystematic and uncritical religious practices of primitive peoples are transformed into the more specifically focused, more regularized, less comprehensively authoritative practices of the more advanced civilizations. Weber, in whom both intellectualism and optimism were rather severely tempered by a chronic apprehensiveness, called this transformation the "disenchantment (*Entzauberung*) of the world." [See DURKHEIM; FRAZER; WEBER, MAX.]

On the heels of evolutionism came, of course, antievolutionism. This took two quite different forms. On one side there was a defense, mainly by Roman Catholic scholars, of the so-called degradation theory. According to this theory, the original revelation of a high god to primitive peoples was later corrupted by human frailty into the idol worship of present-day tribal peoples. On the other side there was an attack, mainly by American scholars of the Boas school, upon the "armchair speculation" of evolutionary thinkers and a call for its replacement by more phenomenological approaches to the study of tribal custom.

The first of these reactions led, logically enough, to a search among the most primitive of existing peoples for traces of belief in a supreme being. The resulting dispute, protracted, often bitter, and stubbornly inconclusive as to the existence of such "primitive monotheism," turned out to be unproductive—aside from some interesting discussions by

Lang (1898) concerning culture heroes and by Eliade (1949) concerning sky gods—and both the issue and the theory that gave rise to it have now receded from the center of scholarly attention. The second reaction has had a longer life and great impact on ethnographic methodology, but it too is now in partial eclipse. Its main contributions—aside from some devastating empirical demolitions of evolutionist generalization—came in the field of cultural diffusion. Leslie Spier's study of the spread of the Sun Dance through the Great Plains (1921) and A. L. Kroeber's application of the age-area approach to aboriginal religion in California are good examples of productive diffusion studies. However, apart from their importance for culture history, the contribution of such distributional studies to our understanding of religious ideas, attitudes, and practices as such has not been great, and few students now pursue these studies. The call of the Boas school for thorough field research and disciplined inductive analysis has been heeded; but its fruits, insofar as religious studies are concerned, have been reaped by others less inhibited theoretically.

### Psychological approaches

The major reaction against the intellectual tradition of the cultural evolutionists took place not within anthropology, however, but in the general context of the positivist revolt against the domination of historicist modes of thought in the social sciences. In the years before World War I the rise of the systematic psychologism of psychoanalysis and of the equally systematic sociologism of the *Année sociologique* forced evolutionist theorizing into the background, even though the leaders of both movements—Freud and Durkheim—were themselves still very strongly influenced by it. Perhaps even more relevant, it introduced a sharp split into anthropological studies of religion which has resolved into the militantly psychodynamic and the militantly social-structural approaches.

Freud's major work in this field is, of course, *Totem and Taboo*, a book anthropologists in general have had great difficulty in evaluating—as Kroeber's two reviews of it, the first facetiously negative, the second, two decades later, ambivalently positive, demonstrate. The source of the difficulty has been an inability or an unwillingness to disentangle Freud's basic thesis—that religious rituals and beliefs are homologous with neurotic symptoms—from the chimerical ethnology and obsolete biology within which he insisted upon setting it. Thus, the easy demolition of what Kroeber called Freud's "just so story" concerning primal incest,

parricide, and guilt within some protohuman horde ("in the beginning was the deed") was all too often mistaken for total rejection of the rather more penetrating proposition that the obsessions, dreams, and fantasies of collective life spring from the same intrapsychic sources as do those of the isolated individual.

For those who read further in Freud's writings, however—especially in "Mourning and Melancholia" and "Obsessive Acts and Religious Practices"—it became apparent that what was at issue was the applicability of theories concerning the forms and causes of individual psychopathology to the explanation of the forms and causes of public myth and group ritual. Róheim (1950) analyzed Australian circumcision rites against the background of orthodox Freudian theories of psychosexual development, especially those clustered around the Oedipal predicament. However, he explicitly avoided recourse to speculations about buried memories of primordial occurrences. Bettelheim (1954) adopted a similar, though more systematic and less orthodox, approach to initiation practices generally, seeing them as socially instituted symbolic mechanisms for the definition and stabilization of sexual identity. Kardiner (1945), taking a neo-Freudian position, sought to demonstrate that the religious institutions of tribal peoples were projections of a "basic personality structure," formed not by the action of an unconsciously remembered historical trauma but by the more observable traumas produced by child-training practices, an approach later extended and cast into quantitative form by Whiting (Whiting & Child 1953). Erikson (1950), drawing upon developments in ego psychology which conceived the emergence of the adult personality to be a joint product of psychobiological maturation, cultural context, and historical experience, interpreted the religious notions of the Yurok and the Sioux in terms of certain basic modes of relating to the world. These relationships gradually developed during the whole course of childhood and adolescence. Others—notably Devereux (1951)—have attempted to use the autobiographical, case-history approach to determine the relations between personality dynamics and religious orientation in particular individuals; still others—notably Hallowell (1937–1954)—have employed projective tests, questionnaires, reports of dreams, or systematic interviews toward similar ends.

In all such studies, even when individual authors have dissented from many of Freud's specific views, the basic premise has been Freudian: that religious practices can be usefully interpreted as expressions

of unconscious psychological forces—and this has become, amid much polemic, an established tradition of inquiry. In recent years, however, responsible work of this type has come to question the degree to which one is justified in subjecting historically created and socially institutionalized cultural forms to a system of analysis founded on the treatment of the mental illnesses of individuals. For this reason, the future of this approach depends perhaps more upon developments within psychoanalysis, now in a somewhat uncertain state, than within anthropology. So far, perhaps only Kluckhohn's pioneering *Navaho Witchcraft* (1944) has attempted to systematically relate psychological factors to social and cultural aspects of primitive religion. The great majority of psychoanalytic studies of tribal beliefs and rites remain willfully parochial.

In any case, not all psychological approaches to religion have been Freudian. Jungian influences have had a certain impact, especially on studies of myth. Campbell (1949), for example, has stressed the continuity of certain themes both cross-culturally and temporally. These themes have been interpreted as expressions of transpersonal constancies in unconscious mental functioning which are at the same time expressions of fundamental cosmic realities.

Simple emotionalist theories have also been extremely popular. There have been two main varieties of these: awe theories and confidence theories. Awe theories have been based on some usually rather vague notion of "religious thrill" experienced by human beings when brought face to face with cosmic forces. A wide range of ethnologists, from Max Müller through Lang and Marett to Lowie and Goldenweiser, have accepted such theories in one form or another. However, awe theories remain mere notations of the obvious—that religious experience is, in the nature of the case, touched with intense feelings of the grandeur of the universe in relation to the self and of the vulnerability of the self in relation to the universe. This is not explanation, but circular reasoning.

Confidence theories also begin with a notion of man's inward sense of weakness, and especially of his fears—of disease, of death, of ill fortune of all kinds—and they see religious practices as designed to quiet such fears, either by explaining them away, as in doctrines of the afterlife, or by claiming to link the individual to external sources of strength, as in prayer. The best-known confidence theory was that set forth by Malinowski. He regarded magic as enabling man to pursue uncertain but essential endeavors by assuring him of



their ultimate success. Confidence, or anxiety-reduction, theories, like awe theories, clearly have empirical foundation but do not adequately explore the complex relationship between fear and religious activity. They are not rooted in any systematic conceptualization of mental functioning and so merely point to matters desperately in need of clarification, without in fact clarifying them.

### Sociological approaches

The sociological approach to the analysis of the religions of nonliterate peoples proceeded independent of, and even at variance with, the psychoanalytic approach, but it shared a concern with the same phenomenon: the peculiar "otherness," the extraordinary, momentous, "set apart" quality of sacred (or "taboo") acts and objects, as contrasted with the profane. The intense aura of high seriousness was traced by Freud to the projection of unacceptable wishes repressed from consciousness onto external objects. The dramatic ambivalence of the sacred—its paradoxical unification of the commanded and the forbidden, the pure and the polluted, the salutary and the dangerous—was a symbolic expression of the underlying ambivalence of human desires. For Durkheim too the extraordinary atmosphere surrounding sacred acts and objects was symbolic of a hidden reality but a social, not a psychological one: the moral force of the human community.

Durkheim believed that the integrity of the social order was the primary requisite for human survival and the means by which that integrity superseded individual egocentricity was the primary problem of sociological analysis. He saw Australian totemism, which he like Freud made the empirical basis of his work, as a mechanism to this end. For example, the collective rituals involving the emblems of the totemic beings—the so-called *totem feasts*—aroused the heightened emotions of mass behavior and evoked a deep sense of moral identification among the participants. The creation of social solidarity was the result of the creation of public veneration by specific groups of persons of certain carefully designated symbolic objects. These objects had no intrinsic value except as perceptible representations of the social identity of the individuals. Collective worship of consecrated trees of painted wood or stone created a *social community*, a "church," upon which rested the vitality of the major social units. These sanctified objects thus represented the system of rights and obligations implicit in the social order and the individuals' unfettered sense of its overriding significance in his life. All sacred objects, beliefs,

and acts, and the extraordinary emotions attending them, were outward expressions of inward social necessities, and, in a famous phrase, God was the "symbol of society." Few anthropologists have been able to swallow Durkheim's thesis whole, when put this baldly. But the more moderate proposition that religious rituals and beliefs both reflect and act to support the moral framework underlying social arrangements (and are in turn animated by it) has given rise to what has become perhaps the most popular form of analysis in the anthropological study of religion. Usually called "functionalism"—or sometimes, to distinguish it from certain variants deemed objectionable, "structuralism"—this approach found its champion in Radcliffe-Brown and its major development in Great Britain, though its influence has now spread very much more widely.

Radcliffe-Brown (1952) agreed with Durkheim's postulate that the main role (or "function") of religion was to celebrate and sustain the norms upon which the integration of society depends. But, unlike Durkheim (and like Freud), Radcliffe-Brown was concerned with the content of sacred symbols, and particularly with the reasons why one object rather than another was absorbed into rite or woven into myth. Why here stones, there water holes, here camp circles, there personified winds?

Durkheim had held this to be an arbitrary matter, contingent upon historical accident or psychological proclivity, beyond the reach of and irrelevant to sociological analysis. Radcliffe-Brown considered, however, that man's need for a concrete expression of social solidarity was not sufficient explanation of the structure of a people's religious system. Something was needed to tie the particular objects awarded sacred status (or, in his terminology, "ritual value") to the particular social interests they presumably served and reflected. Radcliffe-Brown, resolute empiricist that he was, chose a solution Durkheim had already magisterially demolished—the utilitarian. The objects selected for religious veneration by a given people were either directly or indirectly connected to factors critical to their collective well-being. Things that had real, that is, practical, "social value" were elevated to having spiritual, or symbolic, "ritual value," thus fusing the social and the natural into one overarching order. For primitives at least (and Radcliffe-Brown attempted to establish his theory with regard to the sanctified turtles and palm leaves of the preagricultural Andaman Islanders and, later on, with regard to Australian totemism), there is no discontinuity, no difference even, between moral and physical, spiritual and practical relationships

and processes. These people regard both men and things as parts of a single normative system. Within that system those elements which are critical to its effective functioning (or, sometimes, phenomena empirically associated with such elements, such as the Andaman cicada cycle and the shifting monsoons) are made the objects of that special sort of respect and attention which we call religious but which the people themselves regard as merely prudential.

Radcliffe-Brown focused upon the content of sacred symbols and emphasized the relation between conceptions of the moral order of existence and conceptions of its natural order. However, the claim that the sanctity of religious objects derives from their practical social importance is one of those theories which works when it works and doesn't when it doesn't. Not only has it proved impossible to find even an indirect practical significance in most of the enormous variety of things tribal peoples have regarded as sacred (certain Australian tribes worship vomit), but the view that religious concerns are mere ritualizations of real-life concerns leaves the phenomenon of sacredness itself—its aura of mystery, power, fascination—totally unexplained.

More recent structuralist studies have tended to evade both these questions and to concentrate on the role played by religion in maintaining social equilibrium. They attempt to show how given sets of religious practices (ancestor worship, animal sacrifice, witchcraft and sorcery, regeneration rites) do in fact express and reinforce the moral values underlying crucial processes (lineage segmentation, marriage, conflict adjudication, political succession) in the particular society under investigation. Arnold van Gennep's study of crisis rites was perhaps the most important forerunner of the many analyses of this type. Although valuable in their own right as ethnography and as sociology, these structural formulations have been severely limited by their rigid avoidance, on the one side, of the kind of psychological considerations that could account for the peculiar emotions which permeate religious belief and practice, and, on the other, of the philosophical considerations that could render their equally peculiar content intelligible. [See GENNEP.]

### The analysis of symbolic forms

In contrast to other approaches—evolutionary, psychological, sociological—the field of what we may loosely call “semantic studies” of religion is extremely jumbled. There is, as yet, no well-established central trend of analysis, no central

figure around whom to order debate, and no readily apparent system of interconnections relating the various competing trends to one another.

Perhaps the most straightforward strategy—certainly the most disarming—is merely to accept the myriad expressions of the sacred in primitive societies, to consider them as actual ingressions of the divine into the world, and to trace the forms these expressions have taken across the earth and through time. The result would be a sort of natural history of revelation, whose aim would be to isolate the major classes of religious phenomena considered as authentic manifestations of the sacred—what Eliade, the chief proponent of this approach, calls hierophanies—and to trace the rise, dominance, decline, and disappearance of these classes within the changing contexts of human life. The meaning of religious activity, the burden of its content, is discovered through a meticulous, wholly inductive investigation of the natural modalities of such behavior (sun worship, water symbolism, fertility cults, renewal myths, etc.) and of the vicissitudes these modalities undergo when projected, like the Son of God himself, into the flux of history.

Metaphysical questions (here uncommonly obtrusive) aside, the weaknesses of this approach derive from the same source as its strengths: a drastic limiting of the interpretations of religion to the sort that a resolutely Baconian methodology can produce. On the one hand, this approach has led, especially in the case of a scholar as erudite and indefatigable as Eliade, to the uncovering of some highly suggestive clusterings of certain religious patterns with particular historical conditions—for example, the frequent association of sun worship, activist conceptions of divine power, cultic veneration of deified heroes, elitist doctrines of political sovereignty, and imperialist ideologies of national expansion. But, on the other hand, it has placed beyond the range of scientific analysis everything but the history and morphology of the phenomenal forms of religious expression. The study of tribal beliefs and practices is reduced to a kind of cultural paleontology whose sole aim is the reconstruction, from scattered and corrupted fragments, of the “mental universe of archaic man.”

**Primitive thought.** Other scholars who are interested in the meaningful content of primitive religion but who are incapable of so thoroughgoing a suspension of disbelief as Eliade, or are repelled by the cultic overtones of this somewhat mystagogic line of thought, have directed their attention instead toward logical and epistemological considerations. This has produced a long series of studies



that view "primitive thought" as a distinctive mode of reasoning and/or a special body of knowledge. From Lévy-Bruhl through Lévi-Strauss, and with important contributions from members of the evolutionary, psychoanalytic, and sociological schools as well, this line of exploration has persisted as a minor theme in anthropological studies of religion. With the recent advances in linguistics, information theory, the analysis of cognition, semantic philosophy, modern logic, and certain sorts of literary investigation, the systematic study of symbolic activity bids fair to become, in a rather thoroughly revised form, the major theme for investigation. The "new key" Susanne K. Langer heard being struck in philosophy in the early 1940s—"the concern with the concept of meaning in all its forms"—has, like the historicist and positivist "keys" before it, begun to have its echo in the anthropological study of religion. Anthropologists are increasingly interested in ideational expression, increasingly concerned with the vehicles, processes, and practical applications of human conceptualization.

The development of this approach has come in two fairly distinct phases, one before and one after World War II. In the first phase there was a concern with "the mind of primitive man" and in particular with its capacity for rational thought. In a sense, this concern represented the evolutionists' interest in primitive reasoning processes detached from the historicist context. In the second phase, which is still in process, there has been a move away from, and in part a reaction against, the subjectivist emphasis of the earlier work. Ideational expression is thought of as a public activity, rather like speech, and the structure of the symbolic materials, the "language," in whose terms the activity is conducted becomes the subject of investigation.

The first, subjectivist, phase was animated by a protracted wrangle between those who used the religious beliefs and practices of tribal peoples as evidence to prove that there was a qualitative difference between the thought processes of primitives and those of civilized men and the anthropologists who considered such religious activity as evidence for the lack of any such differences. The great protagonist of the first school was the French philosopher Lévy-Bruhl, whose theories of "prelogical mentality" were as controversial within anthropology as they were popular outside it. According to Lévy-Bruhl, the thought of primitives, as reflected in their religious ideas, is not governed by the immanent laws of Aristotelian logical reasoning, but by affectivity—by the vagrant flow of emotion and the dialectical principles of "mystical

participation" and "mystical exclusion." [See LÉVY-BRUHL.]

The two most effective antagonists of Lévy-Bruhl's theories concerning primitive religion were Radin and Malinowski. Radin, influenced by Boas' more general attacks on theories of "primitive mentality," sought to demonstrate that primitive religious thought reaches, on occasion, very high levels of logical articulation and philosophical sophistication and that tribal society contains, alongside the common run of unreflective doers ("men of action"), contemplative intellectuals ("men of thought") of boldness, subtlety, and originality. Malinowski attacked the problem on an even broader front. Using his ethnographic knowledge of the Trobriand Islanders, Malinowski argued that alongside their religious and magical notions (which he, too, regarded as mainly emotionally determined) the "savages" also had a rather well developed and, as far as it went, accurate empirical knowledge of gardening, navigation, housebuilding, canoe construction, and other useful arts. He further claimed that they were absolutely clear as to the distinction between these two sorts of reasoning, between mystical-magical and empirical-pragmatic thinking, and never confused them in actual practice. Of these two arguments, the former seems to be today nearly universally accepted and was perhaps never in fact really questioned. But with respect to the latter, serious doubts have arisen concerning whether the lines between "science," "magic," and "religion" are as simple and clear-cut in the minds of tribal peoples (or any peoples) as Malinowski, never one for shaded judgments, portrayed them. Nevertheless, between them, Radin and Malinowski rather definitively demolished the notion of a radical qualitative gap between the thought processes of primitive and civilized men. Indeed, toward the end of his life even Lévy-Bruhl admitted that his arguments had been badly cast and might better have been phrased in terms of different modes of thinking common to all men. (In fact, Freud, with his contrast between primary and secondary thinking processes, had already made this distinction.) [See MALINOWSKI; RADIN.]

Thus, the debate about what does or does not go on in the heads of savages exhausted itself in generalities, and recent writers have turned to a concern with the symbolic forms, the conceptual resources, in terms of which primitives (and non-primitives) think. The major figure in this work has been Claude Lévi-Strauss, although this line of attack dates back to Durkheim and Mauss's influential 1903 essay in sociological Kantianism, *Primitive Classification*. The writings of E. E.

Evans-Pritchard on Zande witchcraft, Benjamin Whorf on Hopi semantics, and Gregory Bateson on Iatmul ritual and, among nonanthropologists, works by Granet, Cassirer, and Piaget have directed attention to the study of symbolic formulation.

**Symbolic systems.** Lévi-Strauss, whose rather highly wrought work is still very much in progress, is concerned with the systems of classification, the "homemade" taxonomies, employed by tribal peoples to order the objects and events of their world (see Lévi-Strauss 1958; 1962). In this, he follows in the footsteps of Durkheim and Mauss. But rather than looking, as they did, to social forms for the origins and explanations of such categorical systems, he looks to the symbolic structures in terms of which they are formulated, expressed, and applied. Myth and, in a slightly different way, rite are systems of signs that fix and organize abstract conceptual relationships in terms of concrete images and thus make speculative thought possible. They permit the construction of a "science of the concrete"—the intellectual comprehension of the sensible world in terms of sensible phenomena—which is no less rational, no less logical, no more affect-driven than the abstract science of the modern world. The objects rendered sacred are selected not because of their utilitarian qualities, nor because they are projections of repressed emotions, nor yet because they reflect the moral force of social organization ritualistically impressed upon the mind. Rather, they are selected because they permit the embodiment of general ideas in terms of the immediately perceptible realities—the turtles, trees, springs, and caves—of everyday experience; not, as Lévi-Strauss says, apropos of Radcliffe-Brown's view of totems, because they are "good to eat," but because they are "good to think."

This "goodness" exists inherently in sacred objects because they provide the raw materials for analogical reasoning. The relationships perceived among certain classes of natural objects or events can be analogized, taken as models of relationships—physical, social, psychological, or moral—obtaining between persons, groups, or other natural objects and events. Thus, for example, the natural distinctions perceived among totemic beings, their species differentiation, can serve as a conceptual framework for the comprehension, expression, and communication of social distinctions among exogamous clans—their structural differentiation. Thus, the sharp contrast between the wet and dry seasons (and the radical zoological and botanical changes associated with it) in certain regions of Australia is employed in the mythology of the native peoples. They have woven an elaborate origin myth around

this natural phenomenon, one that involves a rain-making python who drowned some incestuous sisters and their children because the women polluted his water hole with menstrual blood. This model expresses and economizes the contrasts between moral purity and impurity, maleness and femaleness, social superiority and inferiority, fertilizing agent (rain) and that which is fertilized (land), and even the distinction between "high" (initiate) and "low" (noninitiate) levels of cultural achievement.

Lévi-Strauss contends that primitive religious systems are, like all symbolic systems, fundamentally communications systems. They are carriers of information in the technical Shannon-Weaver sense, and as such, the theory of information can be applied to them with the same validity as when applied to any physical systems, mechanical or biological, in which the transfer of information plays a central regulative role. Primitives, as all men, are quintessentially multichanneled emitters and receivers of messages. It is merely in the nature of the code they employ—one resting on analogies between "natural" and "cultural" distinctions and relationships—that they differ from ourselves. Where there is a distinguishing difference, it lies in the technically specialized codes of modern abstract thought, in which semantic properties are radically and deliberately severed from physical ones. Religion, primitive or modern, can be understood only as an integrated system of thought, logically sound, epistemologically valid, and as flourishing in France as in Tahiti.

It is far too early to evaluate Lévi-Strauss's work with any assurance. It is frankly incomplete and explorative, and some parts of it (the celebration of information theory, for example) are wholly programmatic. But in focusing on symbol systems as conceptual models of social or other sorts of reality, he has clearly introduced into the anthropology of religion a line of inquiry which, having already become common in modern thought generally, can hardly fail to be productive when applied to tribal myth and ritual.

Whether his own particular formulation of this approach will prove to be the most enduring remains, however, rather more of a question. His rejection of emotional considerations and his neglect of normative or social factors in favor of an extreme intellectualism which cerebralizes religion and tends to reduce it yet again to a kind of undeveloped (or, as he puts it, "undomesticated") science are questionable. His nearly exclusive stress on those intellectual processes involved in classification, i.e., on taxonomic modes of thought (a



reflex of his equally great reliance on totemic ideas as type cases of primitive beliefs), at the expense of other, perhaps more common, and certainly more powerful styles of reasoning, is also doubtful. His conception of the critical process of symbolic formulation itself remains almost entirely undeveloped—hardly more than a sort of associationism dressed up with some concepts from modern linguistics. Partly as a result of this weakness and partly as a result of a tendency to consider symbol systems as entities functioning independently of the contextual factor, many of his specific interpretations of particular myths and rites seem as strained, arbitrary, and oversystematized as those of the most undisciplined psychoanalyst.

But, for all this, Lévi-Strauss has without doubt opened a vast territory for research and begun to explore it with theoretical brilliance and profound scholarship. And he is not alone. As the recent work of such diverse students as Evans-Pritchard, R. G. Lienhardt, W. E. H. Stanner, Victor W. Turner, Germaine Dieterlen, Meyer Fortes, Edmund R. Leach, Charles O. Frake, Rodney Needham, and Susanne K. Langer demonstrates, the analysis of symbolic forms is becoming a major tradition in the study of primitive religion—in fact, of religion in general. Each of these writers has a somewhat different approach. But all seem to share the conviction that an attempt must be made to approach primitive religions for what they are: systems of ideas about the ultimate shape and substance of reality.

Whatever else religion does, it relates a view of the ultimate nature of reality to a set of ideas of how man is well advised, even obligated, to live. Religion tunes human actions to a view of the cosmic order and projects images of cosmic order onto the plane of human existence. In religious belief and practice a people's style of life, what Clyde Kluckhohn called their *design for living*, is rendered intellectually reasonable; it is shown to represent a way of life ideally adapted to the world "as it really" ('fundamentally,' 'ultimately') is." At the same time, the supposed basic structure of reality is rendered emotionally convincing because it is presented as an actual state of affairs uniquely accommodated to such a way of life and permitting it to flourish. Thus do received beliefs, essentially metaphysical, and established norms, essentially moral, confirm and support one another.

It is this mutual confirmation that religious symbols express and celebrate and that any scientific analysis of religion must somehow contrive to explain and clarify. In the development of such an analysis historical, psychological, sociological, and

what has been called here semantic considerations are all necessary, but none is sufficient. A mature theory of religion will consist of an integration of them all into a conceptual system whose exact form remains to be discovered.

CLIFFORD GEERTZ

#### BIBLIOGRAPHY

- BETTELHEIM, BRUNO 1954 *Symbolic Wounds: Puberty Rites and the Envious Male*. Glencoe, Ill.: Free Press.
- CAMPBELL, JOSEPH 1949 *The Hero With a Thousand Faces*. New York: Pantheon.
- DEVEREUX, GEORGE 1951 *Reality and Dream: Psychotherapy of a Plains Indian*. New York: International Universities Press.
- ELIADE, MIRCEA (1949) 1958 *Patterns in Comparative Religion*. New York: Sheed & Ward. → First published as *Traité d'histoire des religions*.
- ERIKSON, ERIK H. (1950) 1964 *Childhood and Society*. 2d ed., rev. & enl. New York: Norton.
- HALLOWELL, A. IRVING (1937-1954) 1955 *Culture and Experience*. Philadelphia: Univ. of Pennsylvania Press.
- KARDINER, ABRAM 1945 *The Psychological Frontiers of Society*. New York: Columbia Univ. Press. → A paperback edition was published in 1963.
- KLUCKHOHN, CLYDE 1944 *Navaho Witchcraft*. Harvard University, Peabody Museum of American Archaeology and Ethnology, Papers, Vol. 22, No. 2. Cambridge, Mass.: The Museum.
- LANG, ANDREW (1898) 1900 *The Making of Religion*. 2d ed. New York: Longmans.
- LESSA, WILLIAM A.; and VOGT, EVON Z. (editors) (1958) 1965 *Reader in Comparative Religion: An Anthropological Approach*. 2d ed. New York: Harper. → Contains a comprehensive bibliography.
- LÉVI-STRAUSS, CLAUDE (1958) 1963 *Structural Anthropology*. New York: Basic Books. → First published in French. See especially the chapters on myth and religion, pages 167-245.
- LÉVI-STRAUSS, CLAUDE (1962) 1966 *The Savage Mind*. Univ. of Chicago Press. → First published in French.
- RADCLIFFE-BROWN, A. R. (1952) 1961 *Structure and Function in Primitive Society: Essays and Addresses*. London: Cohen & West; Glencoe, Ill.: Free Press.
- RÓHEIM, GÉZA 1950 *Psychoanalysis and Anthropology: Culture, Personality and the Unconscious*. New York: International Universities Press.
- SPIER, LESLIE 1921 *The Sun Dance of the Plains Indians: Its Development and Diffusion*. American Museum of Natural History, Anthropological Papers, Vol. 16, part 7. New York: The Museum.
- WHITING, JOHN W.; and CHILD, IRVIN L. 1953 *Child Training and Personality: A Cross-cultural Study*. New Haven: Yale Univ. Press. → A paperback edition was published in 1962.

#### II

#### THE SOCIOLOGY OF RELIGION

Sociologists have undertaken three main types of religious study. They have studied religion as a central theoretical problem in the understanding of social action. They have studied the relation between religion and other areas of social life, such as economics, politics, and social class. And finally,

they have studied religious roles, organizations, and movements. This article will be concerned primarily with the theoretical study of religion and secondarily with the relation between religion and the social structure.

### Historical background

The sociological study of religion has grown out of and remains inextricably related to the much broader effort to understand the phenomenon of religion that has been made by scholars in many fields, especially since the eighteenth century in the West and more recently in other parts of the world. Theologians, philosophers, historians, philologists, literary critics, political scientists, anthropologists, and psychologists have all made contributions. In untangling this immensely complicated story it will be helpful to resort to a simplified schema that focuses on two main lines of intellectual development: the "rationalist" and the "nonrationalist" traditions. (A writer is referred to here as "nonrationalist" not because his thought is considered to be irrational but because he takes the nonrational aspect of human existence as central and irreducible.) Both traditions have roots deep in the history of Western thought and analogues in the thought of some non-Western cultures. The eighteenth century saw a certain crystallization of both traditions, which had important consequences for the nineteenth century and which still affects us in many respects.

**The rationalists and nonrationalists.** The rationalist tradition, which was closely associated with the rise of secular thought and skepticism in England and France in the seventeenth and eighteenth centuries, later stimulated the nonrationalist tendency as a reaction to it. Of course, the rationalists themselves were partly grappling with the great seventeenth-century antirationalist Blaise Pascal. It was the position of the rationalists that the apparently nonrational could not only be brought within the bounds of rational analysis but also could be eliminated as an influence on human action, though most of them doubted that this would ever be possible for the masses. Thus, they held that the obviously absurd doctrines of established religion had come into existence because of ignorance and the deliberate chicanery of the priests, who were serving their own self-interest as well as that of the secular despots who were often their masters. Behind the absurd historical forms, the deists discerned a natural religion in accordance with the dictates of reason.

Rousseau and Kant were transitional figures. Both believed in a generalized "reasonable" religion

in preference to any historical faith, but they grounded their religious convictions on human nature (Rousseau) or the dictates of ethical experience (Kant) rather than on purely cognitive arguments. The nonrationalist tradition, which subsequently developed mainly in Germany, emphasized the *sui generis* quality of religion. Johann Gottfried Herder held that religion was grounded in specific experience and feeling rather than in reason. In the early nineteenth century Friedrich Schleiermacher undertook the first systematic exposition of this position. He held that religion was to be understood neither as crude philosophy nor as primitive ethics but rather as a reality in its own right; it is grounded neither in knowledge nor in action but in feeling. More specifically, he came to think that religion derives fundamentally from the feeling of absolute dependence. Both Herder and Schleiermacher, rejecting as they did a rationalist understanding of religion, opposed the search for a universal natural religion of self-evident reasonableness. Rather, they insisted on taking seriously the particular forms of culture and religion in all their historical diversity. Herder was, of course, one of the most important forerunners of nineteenth-century historicism, and Schleiermacher was one of the first to develop the tools of cultural interpretation (hermeneutics), which provided the main methodological equipment of the movement in Germany.

While the nonrationalist treatment of religion in Germany was closely related to the development of idealist philosophy there, the rationalist treatment of it was closely related to the rise of positivism in France and utilitarianism in England during the nineteenth century. Historicism cut across these distinctions and came to dominate Anglo-French as well as German thought. In both France and England rationalism in its historicist phase took the form of evolutionism.

Comte's famous theory of the three stages viewed theology as appropriate in the childhood of man; however, it was to be displaced first by philosophy and then by science as men's rational understanding of the universe gradually increased. Comte interpreted religion in terms that we would now call "functional," stressing the contribution of belief and ritual to social solidarity and the control of personal feelings. While thus recognizing nonrational factors in religion and, indeed, arguing from them that religion is a permanent and inescapable aspect of human existence, he nonetheless stressed cognitive factors almost exclusively in his theory of religious evolution.

In England, Spencer developed an evolutionary



perspective on religion that was even more narrowly cognitive than Comte's, and Tylor also undertook an extensive effort to understand the religious development of mankind in these terms. In treating the development from animism through polytheism to monotheism as a succession of more and more adequate cognitive hypotheses, he remained thoroughly in the rationalist tradition. Frazer, while formally maintaining the same point of view, embraced a range of material—ritual kingship, human sacrifice, fertility ritual, and the like—which actually demanded a different interpretation. His *Golden Bough* (1890) can be viewed as a marvelously intuitive catalogue of central problems that would have to be solved by other means. Finally, in the English tradition, Marett's discussion (1900) of mana and preanimism came very near breaking through the preconceptions of the rationalist utilitarian school.

Dilthey, who was a follower of Schleiermacher as well as a neo-Kantian, continued the German nonrationalist tradition by stressing the irreducible nature of the religious *Weltanschauung* and the necessity for an inner understanding (*Verstehen*) of its particular forms. This tradition led directly into the modern sociological study of religion through the work of Troeltsch and particularly that of Max Weber, who both transcended the tradition in important respects. Postponing a consideration of these men for a moment, we may trace further into the twentieth century the development of the German tradition that was relatively independent of the influence of Weber and Troeltsch. A certain formal culmination was reached in the work of Rudolph Otto. In his important book *The Idea of the Holy* (1917), he richly developed the basic intuition of Schleiermacher by spelling out the phenomenology of the holy in terms of the numinous—the notion of the *mysterium tremendum et fascinosum*. Here more clearly than ever was the assertion of the *sui generis* nature of religion and its "geometrical location" in a certain kind of immediate experience. That experience, according to Otto, can be phenomenologically understood, but it cannot be explained. This point was made quite explicit in the work of van der Leeuw, who shared Otto's essential position: in the preface to *Religion in Essence and Manifestation* van der Leeuw said that he specifically disavowed all "theories" ([1933] 1938, p. vi). Thus Otto and the later exponents of this tradition—whether they were phenomenologists (such as van der Leeuw and Mircea Eliade) or were specifically interested in the sociology of

religion (Joachim Wach and Gustav Mensching)—have given us a rich array of materials without appreciably advancing our theoretical understanding.

Generalizing the argument so far, we may say that while the nonrationalist tradition jealously guarded the specific nature of religion but eschewed any explanation of it, the rationalist tradition provided a number of ways of explaining religion which in the end explained it away. We will argue, following Talcott Parsons' *The Structure of Social Action* (1937), that around the turn of the twentieth century several men from both traditions broke free of their preconceptions and converged on a more adequate approach to religion (and indeed to social phenomena generally).

However, just as we have noted that the non-rationalist tradition continues to our own day in relatively pure form, so we must point out that the utilitarian rationalist position also continues, not so much as a conscious tradition of scholarship as a semiconscious preconception of scholars in many fields. This preconception has been strengthened by its gradual fusion with a Marxian understanding of religion as essentially an ideological cover, either for the defense of the social status quo or for protests against it. Here religion is treated as a weapon in the political-economic struggle or as a preliminary stage in a political movement, a stage that may be outgrown with the attainment of political maturity. What is at issue is not the empirical adequacy of such analyses in particular instances, but rather the generalization of them as adequate for the understanding of the phenomenon of religion.

**Foundations of an adequate theory.** The architects of a more adequate understanding of religion were Durkheim, Weber, and Freud, though others also made important contributions.

**Contributions of Weber and Durkheim.** Weber maintained the idea of the centrality and irreducibility of nonrational elements in human action as it had developed in the German tradition, but he was not content with a mere phenomenological description of these elements. According to Parsons' analysis, Weber began, at least incipiently, to place these nonrational elements within the context of a general theory of social action. This he did through two of his central concerns. The first was with the problems of *meaning*—of evil, suffering, death, and the like—which are inescapable in human life but insoluble in purely scientific terms. Weber argued that the alternative religious answers to these problems of meaning

have had not only profound consequences for the motivation of individuals but also, in the long run, important causal effects on social development. The second of Weber's concerns, which links irrational elements to a more general theory of action, was with what he called *charisma*. Charisma is primarily a quality of the individual that places him above normal expectations and endows him with the authority to utter new commandments. Charisma is a relational concept—that is, it comes into existence only when it is recognized by a group. It links deeper levels of psychic organization, within the charismatic individual and in the members of the group who recognize him, with the social process and particularly with the possibility of radical discontinuities in social development. In both cases Weber was arguing for the importance of religion in social action on the grounds of its closeness to powerful nonrational motivational forces and its capacity to give form and pattern to those forces, including its capacity to create radically new forms and patterns.

The rationalist positivist tradition was decisively broken through from within by Durkheim, when he recognized that religion is a reality *sui generis*. By this he meant that religious representations or symbols are not delusions, nor do they simply stand for some other phenomena, such as natural forces or (contrary to some interpretations of his work) social morphology. Rather, in his social Kantianism, he held that religious representations are constitutive of society. They exist within the minds of individuals so as to inhibit egocentric impulses and to discipline the individual so that he can deal objectively with external reality. These shared representations, with their capacity to direct and control personal motivation, are what make society possible. While his treatment of motivation remained rudimentary, he did point out very clearly the great importance of religious action for stimulating individuals to participate positively in social life (see the discussion of collective effervescence in Durkheim 1912, book 2, chapter 7) and for dealing with tendencies of individuals to withdraw from social life (see the discussion of funeral ritual in 1912, book 3, chapter 5).

Although they started from opposite directions, both Weber and Durkheim seem to have overcome the impasse in which the rationalist and non-rationalist approaches to religion had long been caught: they both placed religion in a theoretical rather than a purely descriptive context, without denying its centrality and irreducibility. However,

they lacked any fine-grained structural understanding of the nonrational elements involved, even though they recognized their importance. It was in the work of Freud that a structural understanding of the relevant emotional and motivational elements was to be found for the first time.

*Freud's work on religious symbolism.* Freud's early work on the stages of psychosexual development was applied to religion in *Totem and Taboo* (1913), where the Oedipus complex with its mixture of dependence, love, and hostility was seen as the core problem of religious symbolism, which he treated as largely projective. Freud's later ego psychology, heralded in the important essays "Group Psychology and the Analysis of the Ego" (1921) and "The Ego and the Id" (1923), provided the basis for a much more active understanding of religious symbolism, which could now be treated not merely as reflecting psychic conflict but as actually affecting the outcome of psychic conflict and redirecting psychic forces. This point of view was applied to religion in the somewhat idiosyncratic but extraordinarily fruitful *Moses and Monotheism* (1934–1938). "The Future of an Illusion" (1927) represents a reversion to the early projective theory of religion and is neither the final statement of Freud's position nor even typical of his own late thinking.

By the early 1920s, then, the elements of a more adequate theory of religion had come into existence. However, at just this point the primary preoccupation with religion displayed by most of the great social scientists of the late nineteenth and early twentieth centuries disappeared. Other issues occupied the center of attention. Even today a theoretical concern with religion is only gradually reviving as a central issue in social science.

### A theory of religion

Only in the last few years has a new model of human action developed that will allow us to utilize the insights of Weber, Durkheim, and Freud without falling back into the old controversies about idealism and materialism, rationalism and irrationalism, and humanism and science. This is the cybernetic model (for a relevant, though partial, exposition, see Deutsch 1963). Parsons' action theory is the chief link between earlier social science and cybernetic theory and has itself in recent years been increasingly restated in cybernetic terms. This model, whose basic terms are "energy" and "information," seems likely to integrate the range of behavioral research



that stretches from the work of biologists such as Nikolaas Tinbergen and J. P. Scott on animal behavior to that of philosophers such as Ernst Cassirer and Susanne K. Langer on symbolic forms.

**The cybernetic model.** An action system may be defined as the symbolically controlled, systematically organized behavior of one or more biological organisms. The energy of such a system is supplied by the organism and is organized through genetically controlled organic structures that are not directly open to symbolic influence. Thus the basic motivation of the action system—its drives and needs—is partially determined by organic structure, although it is subject, through learning processes, to a considerable degree of symbolic control. The precise boundary, or, perhaps better, the area of overlap, between what is genetically and what is symbolically controlled is by no means clear, and important research on the problem is continuing. For present purposes it is only necessary to note the relatively broad plasticity for symbolic learning that is usually recognized, at least in the human species.

Information in such a system consists largely of symbolic messages that indicate something about either the internal state or the external situation of the action system. These messages are understood or interpreted by matching them with previously learned symbolic patterns, which make up the memory of the system. New situations can be understood through a recombination of previously learned elements, so that a new symbolic pattern is created. Thus, the set of symbolic patterns existing in a system will be partially determined by the nature of the external world with which that system has had to deal and by the nature of the laws governing the cognitive processes of the brain. Within these limits there is a wide range of freedom, within which alternative symbolic patterns may operate with equal or nearly equal effectiveness, as is best illustrated perhaps in the phenomenon of language.

The cybernetic model thus conceives of a human action system as autonomous, purposive, and capable of a wide range of external adaptations and internal rearrangements within the very broad constraints inherent in the nature of energy and information. It is precisely the stress on autonomy, learning capacity, decision, and control that gives the cybernetic model the ability, lacking in previous mechanistic and organic models, to assimilate the contributions of the humanistic disciplines—the *Geisteswissenschaften*—without abandoning

an essentially scientific approach [see *SYSTEMS ANALYSIS, articles on SOCIAL SYSTEMS and POLITICAL SYSTEMS*].

The openness of an action system is always relative to its particular structure. We have noted that energy is structured by genetic information from the organism and by symbolic information from the culture. Energy and information are further structured by two systems that can be seen as integrating organic motivation and symbols: personality and society. These are symbolically patterned motivational systems; the first centers on individual organisms, and the other on groups. Any particular personality or social system will be determined in large part at any given moment by its history, for its learning capacity will be largely a product of the structures it has built up over time. However, if a system has managed to develop a broad and flexible capacity for rapid learning, and if it has a reserve of uncommitted resources, its reaction to any given internal or external situation will be open to a wide variety of alternative choices; in other words, it will have a high degree of freedom.

**The role of religion in action systems.** Religion emerges in action systems with respect to two main problems. In order to function effectively, it is essential that a person or group have a relatively condensed, and therefore highly general, definition of its environment and itself. Such a definition of the system and the world to which it is related (in more than a transient sense) is a conception of *identity*. Such a conception is particularly necessary in situations of stress and disturbance, because it can provide the most general set of instructions as to how the system is to maintain itself and repair any damage sustained.

In addition to the identity problem, there is the problem of dealing with inputs of motivation from within the system that are not under the immediate control of conscious decision processes. We have already noted that such motivation is partly under the control of genetic rather than symbolic processes. In addition, as Freud discovered, there are important symbolically organized systems of motivation that are partially blocked or screened off from consciousness through the mechanism of repression; this happens partly as a result of the pattern of child raising. Though these unconscious motivational forces exist in individuals, they are to a large degree shared in groups, because human biology and child-raising patterns are, broadly speaking, similar. The emotions can be seen as signaling devices by which the conscious decision

process (the ego, in Freud's terms) becomes aware of the existence of important inputs of motivation from unconscious levels of the system.

The problems of identity and of unconscious motivation are closely related, for it is just those situations of threat, uncertainty, and breakdown, which raise the identity issue, that also rouse deep unconscious feelings of anxiety, hope, and fear. An identity conception capable of dealing with such a situation must not only be cognitively adequate but must also be motivationally meaningful. It is precisely the role of religion in action systems to provide such a cognitively and motivationally meaningful identity conception or set of identity symbols. It is such a set of symbols that provides answers to Weber's problems of meaning and to which Durkheim referred in speaking of religious representations. It is also what in large part Freud meant when he spoke of the superego. This mode of analysis leads to a definition of religion very close to that of Clifford Geertz, who wrote: "A religion is a system of symbols which acts to establish powerful, pervasive and long-lasting moods and motivations in men by formulating conceptions of a general order of existence and clothing these conceptions with such an aura of factuality that the moods and motivations seem uniquely realistic" (1966, p. 4).

The conception of religion briefly sketched here—religion as the most general mechanism for integrating meaning and motivation in action systems—applies to all types of action systems, not only to whole societies or groups of them. Many smaller units (individual personalities and groups) appropriate the religious symbols of their social and cultural environment in dealing with their own religious problems, though always with some degree of individual variation. Moreover, even where prevailing religious symbol systems are rejected, the idiosyncratic solutions of individuals and groups to fundamental problems of orientation and identity may be viewed in terms of this scheme as "religious." Of course, the degree to which religious problems will be salient for any individual or group is quite variable.

#### Current theoretically relevant research

While the theory outlined above has not previously been stated in exactly these terms, some recent research seems to be guided by a conception of religion as a control system linking meaning and motivation—a conception that is a close parallel of this theory. Three particularly interesting examples of such research will be cited. Each of

them concentrates at a traditional level of social science—psychological, social, or cultural—but it is noteworthy that they all transcend the discipline boundaries in trying to understand religion.

**Analysis of historical figures.** The first is the application of detailed psychological analysis to significant figures in religious history. Erik H. Erikson, in *Young Man Luther* (1958), has overcome a tendency toward psychological reductionism that was evident in some earlier work of this sort. He is careful to place Luther's life history in its proper social, historical, and religious context. In doing so, he shows how Luther was able to solve his own problem of identity, which had its roots deep in his unconscious conflicts, through constructive innovation with respect to religious symbolism. However, Erikson's knowledge of the deep motivational forces involved does not lead him to deny the contribution of the religious reformulation itself. Rather, he emphasizes the fusing and forming power of the symbols to synthesize motivational conflicts and control destructive impulses.

Erikson also contributes to the broader problem of religious change by indicating that Luther's solution, once it was embodied in communicable symbolic form, could be appropriated by others in the same society who had analogous identity problems arising from social-historical matrices similar to Luther's own. Moreover, although the argument is not worked out in detail, Erikson implies that Luther was able to contribute to certain social and cultural identity problems of Germany in particular and of Western civilization in general. This detailed analysis of the mechanism involved in a major example of religious change, although it involves only one among several types of change, is an important contribution to one of our least understood problems. [See **IDENTITY, PSYCHO-SOCIAL.**]

**Small-group research.** The second new line of research has to do with the analysis of religious phenomena in small groups, as illustrated in Philip E. Slater's *Microcosm* (1966). The small-group situation provides a laboratory within which the symbolic processes involved in group formation, as hypothesized by Durkheim and Freud, can be empirically synthesized. Slater shows how the group and its leader become foci for the unconscious anxieties, hopes, and fears of the members in the initial situation of uncertainty and lack of definition. These feelings are handled initially by the development of projective symbols, which are redolent with the mysterious potency of the unconscious feelings, that is, are highly magical and



sacred. However, over time, as more and more of the group process becomes consciously understood and as the deeper feelings about the group and the leader are worked through, the group itself becomes "secularized" and the quality of the sacred is relegated to certain group ideals and symbols of group solidarity.

Slater's comparison of this process of group development with the long-term process of religious evolution is highly suggestive. Particularly important is the perception of the changing role of symbolism with respect to the meaning-motivation balance as the group itself undergoes new experiences and changes in structure. Also of great general significance is Slater's observation that social and psychological aspects of group process interpenetrate at every point and that a purely psychological or a purely sociological theoretical framework would result in an inadequate understanding of the processes at work.

**Analysis of language as symbolic action.** A third area of promising research is the analysis of language as symbolic action as it has been developed by Kenneth Burke in his *Rhetoric of Religion* (1961) and other works. Burke's theory of artistic form (defined as "the arousing and fulfilling of expectations"), which is based on the Aristotelian theory of catharsis, emphasizes the working through of emotional tensions in symbolic form; thus, it is very close to the theory of religion developed above. Burke has pointed out that the tensions may be social as well as psychological, though they must be presented in personal terms if they are to have emotional impact. He has also pointed out the function of social and psychological control that such symbolic working through performs.

Perhaps Burke's chief contribution is his insistence on the special formal qualities of language and symbolism generally. Language does not merely reflect or communicate social and psychological realities. Rather, language contains within itself a "principle of perfection" (1961, epilogue), which operates relatively independently of other factors. That is, a terminology, once established, has certain "entelechia" implications. For example, once the terms *king*, *society*, and *god* come into existence, there is the logical possibility that someone will ask what a perfect king, society, or god is and from this analysis actually draw conclusions about existing kings, societies, or ideas of god. Regardless of the many actual social and psychological blocks to carrying through such analyses, the fact is that the possibility of pushing toward terminological perfection is always

present, and under certain circumstances it will actually be pursued. This is one of the chief reasons why cultural symbol systems cannot be treated entirely reductionistically. In *The Rhetoric of Religion* Burke has applied these insights with great brilliance to two key religious documents: the first three chapters of Genesis and Augustine's *Confessions* [see INTERACTION, article on DRAMATISM].

### Religion and society

An interest in the social consequences of religious belief and action is probably as old as the interest in religion itself. In the sixteenth century Machiavelli gave, in his *Discourses*, a functional analysis of Roman religion which had considerable influence. Spinoza, Montesquieu, Rousseau, and Comte continued and developed the analysis of the political significance of religious commitment and the related problems of the influence of religion on personal morality and social solidarity. In one of the earliest quantitative studies in sociology, Durkheim introduced a "religious variable" in his study of suicide (1897). Subsequently many studies of varying breadth and quality have included a religious variable and thus advanced our knowledge of the influence or lack of influence of religion on some aspect of social existence; however, a comprehensive survey of the results obtained is still lacking.

The man who contributed most to the systematic understanding of the interrelations between religion and society and who stimulated more research (both quantitative and nonquantitative) than any other scholar was undoubtedly Max Weber. The most influential single hypothesis in this field is certainly his thesis on the influence of the Protestant ethic on the rise of modern society (1904-1905). Although the cumulated argument and evidence on this thesis are too massive to review here, one recent study by Gerhard Lenski (1961) is worth mentioning, not only because it sheds light on the Protestant-ethic thesis in contemporary America but also because it is perhaps the most successful attempt to apply the methods of survey research to the sociology of religion. Lenski was concerned with the influence of religious affiliations and beliefs on attitudes toward work, authority, education, and a variety of other matters in a large American city, and he found this influence to be considerable. Weber's ideas on the non-Western religions have been applied much more sporadically (see the incomplete review in Bellah 1963), but as the comparative horizon of the sociology of religion widens, these

ideas seem likely to attract more attention. In this same connection, a revival of interest in religious evolution as providing the soundest classificatory system for comparative work can also be expected.

### Religious evolution

Although foreshadowings of the idea of religious evolution can be traced as far back as classical times, the first extensive effort in this direction was that of Vico in the eighteenth century. Elaborate schemes of religious evolution with copious empirical illustration were developed in the nineteenth century by Hegel, Comte, and Spencer. In more modest and judicious form, evolutionary ideas provided the basis of the sociology of religion of Durkheim (1912) and Weber (1922). Though long neglected and in some quarters excoriated, the idea of religious evolution has recently been revived (Bellah 1964). It provides the natural link between the kind of theory of religion sketched in this article and the comparative study of religion.

If one defines religion as a control system linking meaning and motivation by providing an individual or a group with the most general model that it has of itself and its world, then it becomes apparent that such a control system can vary in degree of complexity in ways that are not entirely random with respect to the degree of complexity of the social system of which it is a part. Since it has been clear for a long time that levels of social and cultural complexity are best understood in an evolutionary framework, it seems inevitable that religion too must be considered in such a framework.

Here it is possible only to enumerate some of the most important dimensions along which religious evolution can be expected to occur. The central focus of religious evolution is the religious symbol system itself. Here the main line of development is from compact to differentiated symbolism, that is, from a situation in which world, self, and society are seen to involve the immediate expression of occult powers to one in which the exercise of religious influence is seen to be more indirect and "rational." This is the process of the "disenchantment of the world" that was described by Weber. Part of this process is the gradual differentiation of art, science, and other cultural systems as separate from religious symbolism. Furthermore, changes in the nature and position of religious symbolism effect changes in the conception of the religious actor. The more differentiated symbol systems make a greater demand on the individual for decision and commitment.

To support this growing religious individualism, specifically religious group structures are required, whereas at earlier stages religion tends to be a dimension of all social groups. Finally, the capacity for religion to provide ideals and models for new lines of social development increases with the growing symbolic, individual, and social differentiation.

An adequate theory of religious evolution would have to go hand in hand with a general theory of social evolution. For example, the varying saliency of law, custom, and religion as agencies of social control in different societies, which was long ago pointed out by Montesquieu, could probably best be explicated by a general analysis of the evolution of control systems. Moreover, the contribution that a theory of religious evolution could make to a general conception of social evolution is also considerable. Such an approach would begin to indicate the ways in which changes in social structure impinge on the integration of established cultural meanings with the deeper levels of individual personalities and how shifts in that meaning-motivation balance can in turn help or hinder social differentiation.

ROBERT N. BELLAH

### BIBLIOGRAPHY

- BELLAH, ROBERT N. 1963 Reflections on the Protestant Ethic Analogy in Asia. *Journal of Social Issues* 19: 52-60.
- BELLAH, ROBERT N. 1964 Religious Evolution. *American Sociological Review* 29:358-374.
- BURKE, KENNETH 1961 *The Rhetoric of Religion: Studies in Logology*. Boston: Beacon.
- DEUTSCH, KARL W. 1963 *The Nerves of Government: Models of Political Communication and Control*. New York: Free Press.
- DURKHEIM, ÉMILE (1897) 1951 *Suicide: A Study in Sociology*. Glencoe, Ill.: Free Press. → First published in French.
- DURKHEIM, ÉMILE (1912) 1954 *The Elementary Forms of the Religious Life*. London: Allen & Unwin; New York: Macmillan. → First published as *Les formes élémentaires de la vie religieuse, le système totémique en Australie*. A paperback edition was published in 1961 by Collier.
- ERIKSON, ERIK H. (1958) 1962 *Young Man Luther: A Study in Psychoanalysis and History*. Austen Riggs Monograph No. 4. New York: Norton.
- FRAZER, JAMES (1890) 1955 *The Golden Bough: A Study in Magic and Religion*. 3d ed., rev. & enl. 13 vols. New York: St. Martins; London: Macmillan. → An abridged edition was published in 1922 and reprinted in 1955.
- FREUD, SIGMUND (1913) 1959 Totem and Taboo. Volume 13, pages ix-162 in *The Standard Edition of the Complete Psychological Works of Sigmund Freud*. London: Hogarth; New York: Macmillan. → First published in German.



- FREUD, SIGMUND (1921) 1955 *Group Psychology and the Analysis of the Ego*. Volume 18, pages 67-143 in *The Standard Edition of the Complete Psychological Works of Sigmund Freud*. London: Hogarth; New York: Macmillan. → First published in German.
- FREUD, SIGMUND (1923) 1961 *The Ego and the Id*. Volume 19, pages 12-63 in *The Standard Edition of the Complete Psychological Works of Sigmund Freud*. London: Hogarth; New York: Macmillan. → First published in German.
- FREUD, SIGMUND (1927) 1961 *The Future of an Illusion*. Volume 21, pages 5-58 in *The Standard Edition of the Complete Psychological Works of Sigmund Freud*. London: Hogarth; New York: Macmillan. → First published in German.
- FREUD, SIGMUND (1934-1938) 1964 *Moses and Monotheism: Three Essays*. Volume 23 in *The Standard Edition of the Complete Psychological Works of Sigmund Freud*. London: Hogarth; New York: Macmillan. → First published in German.
- GEERTZ, CLIFFORD 1966 *Religion as a Cultural System*. Pages 1-46 in *Conference on New Approaches in Social Anthropology*, Jesus College, Cambridge, England, 1963, *Anthropological Approaches to the Study of Religion*. Edited by Michael Banton. A.S.A. Monographs, Vol. 3. London: Tavistock; New York: Praeger.
- LEEUW, GERARDUS VAN DER (1933) 1938 *Religion in Essence and Manifestation: A Study in Phenomenology*. New York: Macmillan. → First published as *Phänomenologie der Religion*. A paperback edition was published in 1963 by Harper.
- LENSKI, GERHARD E. (1961) 1963 *The Religious Factor: A Sociological Study of Religion's Impact on Politics, Economics, and Family Life*. Rev. ed. Garden City, N.Y.: Doubleday.
- LESSA, WILLIAM A.; and VOGT, EVON Z. (editors) (1958) 1965 *Reader in Comparative Religion: An Anthropological Approach*. 2d ed. New York: Harper.
- MARETT, ROBERT R. (1900) 1929 *The Threshold of Religion*. 4th ed. London: Methuen.
- MENSCHING, GUSTAV 1947 *Soziologie der Religion*. Bonn: Röhrscheid.
- OTTO, RUDOLPH (1917) 1950 *The Idea of the Holy: An Inquiry Into the Non-rational Factor in the Idea of the Divine and Its Relation to the Rational*. 2d ed. London: Oxford Univ. Press. → First published as *Das Heilige*.
- PARSONS, TALCOTT (1937) 1949 *The Structure of Social Action: A Study in Social Theory With Special Reference to a Group of Recent European Writers*. Glencoe, Ill.: Free Press.
- SLATER, PHILIP E. 1966 *Microcosm*. New York: Wiley.
- TROELTSCH, ERNST (1912) 1960 *The Social Teaching of the Christian Churches*. 2 vols. New York: Harper. → First published as *Die Soziallehren der christlichen Kirchen und Gruppen*.
- WACH, JOACHIM (1944) 1951 *Sociology of Religion*. Univ. of Chicago Press.
- WEBER, MAX (1904-1905) 1930 *The Protestant Ethic and the Spirit of Capitalism*. Translated by Talcott Parsons, with a foreword by R. H. Tawney. London: Allen & Unwin; New York: Scribner. → First published in German. The 1930 edition has been reprinted frequently.
- WEBER, MAX (1922) 1963 *The Sociology of Religion*. Boston: Beacon. → First published in German as Book 2, Chapter 4 of *Wirtschaft und Gesellschaft*.

### III PSYCHOLOGICAL STUDY

Religion affords dramatic, insistent phenomena centrally entrenched in the intimate structures and functions of personality. These phenomena are likely to intrigue a psychologist for one of two contrasting reasons: either for their challenging uniqueness or for their vivid illustration of factors of more fundamental and general interest in psychology.

Sometimes, such phenomena as sudden conversion, mysticism, glossolalia, ritual, and intense religious commitment may seem incomprehensible in terms of readily available psychological theories and therefore may spark the development of new ones. Alternatively, they may appear to be particularly vivid instances of processes of central importance to psychology. In this case, these phenomena become a challenge in that they demand an account from those who aspire to a comprehensive theory of personality and its development and change, motivation, emotion, cognition, perception, group behavior, attitudes, values, socialization, guilt, or virtually any other theme of common psychological interest. In particular, religion poses urgent questions about the relation between affective and cognitive processes and between the individual and his social environment.

Religion appears to share with abnormal phenomena this characteristic of presenting seemingly dramatic and unique manifestations which at the same time illustrate processes of central concern to any personality psychologist.

This distinction—between the seeming uniqueness of religion and its instancing of general problems in psychology—suggests alternative strategies in the study of the psychology of religion. One strategy begins with careful attention to the phenomena and to the development of useful descriptive categories to handle them. The other strategy turns to religious phenomena only after a theory has been developed in another segment of personality and behavior. Of the pioneer systematists at the beginning of this century who turned their attention to religion, James (1902) represents the former strategy, Freud (1913; 1927; 1934-1938) the latter.

The phenomena-based strategy generally, although of course not exclusively or necessarily, has been employed by those who were personally more "sympathetic" to religion, who therefore tended to claim a greater richness or validity for their "participant" observations.

The psychology of religion has been marked by a schism between these two strategies. Descriptive,

phenomena-based approaches of various kinds have made little progress toward developing genotypical categories useful for theoretical analysis. On the other hand, the "theoreticians" starting "outside" religion have made little use of the distinctions or other descriptive gains made by those interested in the phenomena.

While empirical research by either group has been limited—because practical, design, and measurement problems are formidable—the split, with regard to empirical work, is particularly wide. The theorists have apparently felt that any empirical obligation was already satisfied by demonstration—often clinical—of their theory before they applied it to religion or else by anecdotal application to instances derived from "armchair" observations or library resources. Empirical research by the phenomena-based group has tended to be limited to the most superficial data, such as church attendance or questionnaire self-report of "belief." Questions of greatest interest and importance have not been asked.

**Background and context.** Beyond the general psychological fascination which religion may have, impetus for work in the psychology of religion has come from at least three sources. The first, and perhaps the major, impetus has come from practical and theoretical concerns within institutionalized religion, especially Protestantism.

For example, the field of religious education has provided prominent focus, suggesting both problems and theories. Especially among Protestants in the early decades of this century, as sudden adolescent conversion became less normative, those concerned with religious education required that more attention be paid to the processes of religious development. This was encouraged by an optimistic, "liberal" theology which anticipated (in harmony with other progressive ideologies of the time) the possibility of developing ideal Christians, with the aid of scientifically discoverable principles. More recent "existential" concerns, emphasizing the role of "experience" in Christian nurture, as distinguished from intellectual instruction, has again brought new urgency to psychological questions.

During the middle decades of the twentieth century there has also been concern, especially among Protestant and Jewish clergymen, with finding sound psychological bases for dealing, as pastors, with personal and family crises. The acceptance of "depth psychology" has coincided with the development of a less optimistic "neo-orthodox" theology which emphasizes the darker side of human existence rather than its progressive improvement.

Furthermore, the general loosening of traditional

religious institutions and roles has sometimes left clergymen searching for a clearer pragmatic basis for defining, and perhaps even justifying, their role. This concern has coincided with greater institutional concern with recruiting, screening, training, and evaluating clergymen. The resulting research on the role, personality characteristics, and motivation of clergymen (Menges & Dittes 1965) has represented a new focus for the psychology of religion. [See PSYCHIATRY, *article on THE RELIGIO-PSYCHIATRIC MOVEMENT*.]

A second impetus, beyond the bounds of institutionalized religion, is a single frequently replicated empirical finding in social psychology: the positive correlation between indices of religion and of prejudice, authoritarianism, or other conservative attitudes. This finding has occasioned a continuing sequence of empirical and theoretical work for two decades. This has perhaps been provoked in part by the conflict in values aroused by the finding: religion is commonly regarded as a desirable value, and prejudice is not, and the official doctrines of organized religions generally repudiate prejudice.

The third impetus arises from the intrinsic connections between religion and psychology, especially clinical psychology. In many ways, religious and psychological enterprises have parallel, if not overlapping, concerns. Both call attention to, and offer conceptualizations about, human aspirations and apprehensions, about the relation of an individual to his total environment, about the processes of coping with stress, and about ameliorating the human condition. It is not surprising, then, that each should be intrigued by the prospective insights of the other.

Contemporary theologians (e.g., Roberts 1950; Tillich 1952) have systematically attempted to make theology take account of psychological formulations. Relatively little "adaption" in the other direction—exploitation of theological formulations for psychological research—has been attempted, although it would seem profitable (Hiltner & Rogers 1962). Theological formulations frequently include implicit or explicit psychological theories, including propositions about both the antecedents and the consequences of particular beliefs or other religious behavior.

**Philosophical and logical distinctions.** Some methodological attention must be paid to the important distinctions between religion and psychology. Psychological analysis of religion operates within certain clear and logically necessary, but not always understood, restrictions which distinguish it from other intellectual disciplines related to religion. There is sometimes a special problem



in relating, or distinguishing between, the psychology of religion and both philosophy and theology.

Logically, the distinction seems clear and has been made in similar terms by writers in the psychology of religion; James devoted an eloquent chapter to the point (1902). Psychology, philosophy, and theology may all be concerned with the same phenomenon, such as a particular belief or a particular ritual, but they ask different questions about the phenomenon. Psychological analysis is concerned with the psychological development and the functions of the belief or ritual. Philosophy is concerned with the correspondence between the belief or ritual and some criterion of truth, logic, or goodness, and theology, with its correspondence with such criteria as the will of God or other given norms of faith. The answers to the psychological questions do not necessarily imply or presuppose answers to the philosophical or theological questions. Even a thorough assessment of the psychological history and functions of a particular belief carries no ordinary implications for the "truth" or "faithfulness" of the belief; these must still be ascertained by the criteria appropriate to philosophy and theology. Both "true" and "false" beliefs may develop out of a particular pattern of social influences or personal motivation. The "genetic fallacy"—the evaluation of belief or behavior in terms of an analysis of its origins—has been soundly discounted by every major writer in the psychology of religion, including Freud (1927), who carefully used the term "illusion," when applied to religion, to refer to certain wish-fulfilling functions, not to religion's objective validity. Freud even acknowledged that science, a preference for which he argued on other grounds, was an "illusion" in this same sense.

Yet, despite the logical distinction—well accepted in principle—between psychology, on the one hand, and philosophy and theology, on the other, there remain at least two psychologically understandable grounds for the continuing confusion and blurring—perhaps more among the audience of the psychologist than by the psychologist himself.

One of these grounds for confusion is an unfamiliarity with, or a mistrust of, discrete disciplines of thought which claim correspondence, not with the "whole truth," but only with an abstracted and limited segment. Most socializing and reality-orienting processes accustom persons to regard reality as unitary and therefore to expect statements about reality to be similarly unified. Discrete disciplines or systems of thought are expected to be integrated. Differing statements—for example,

psychological and philosophical statements about the same belief—are perceived as mutually implicative. If mutual implications cannot be recognized, the statements are perceived as necessarily inconsistent.

The attempt to integrate different statements into a single statement or system often leads to an unwarranted syncretistic identification of terms (for example, "neurosis" and "sin" or "psychological anxiety" and "ontological anxiety") drawn from discrete psychological and theological systems (e.g., Roberts 1950; Tillich 1952). When such integration is not easily accomplished, then the psychological and philosophical (or theological) systems are seen as antagonistic, competing for the status of the *one* valid approach. With respect to Copernican astronomy, geology, and biological evolution, the result of long battles is the general acknowledgment that statements about history and function and statements about meaning and truth may *both* be true without having to be consistently integrated. But this acceptance is not yet universally accorded to psychological analysis. The view commonly persists that a belief must be either psychologically motivated *or* true (e.g., Havens 1961; Outler 1954).

A second psychological basis for confusion is a persistent notion in our culture, apparently related to Greek philosophy and perhaps to America's Puritan roots, that psychological history and motives are "unworthy" vehicles for truth. Epistemologically and theologically, it may be argued, in principle, that truth may readily arise through such psychologically discernible processes as social or motivational influences, as well as through pure reasoning; such a view would seem especially consistent with Christian notions of incarnation or Jewish emphasis on the importance of the community. But psychologically, the distinction between rational and emotional aspects of man doggedly persists in our culture and, hence, in psychologies of religion (e.g., Allport 1950; Clark 1958), with higher status accorded the rational aspects. Thus, a discipline, such as philosophy, that is concerned with intellectual processes and criteria may be accorded a superior role, to which the contributions of psychology must be accommodated [e.g., Outler 1954; see also KNOWLEDGE, SOCIOLOGY OF].

**Phenomena-based approaches.** The search within the phenomena of religion for useful descriptive and analytical categories has, ironically, been impeded by the dramatic strikingness of the phenomena. It has been found expedient to conceptualize religion largely in terms of particular phenomena: for example, conversion experience, church at-

tendance, personal prayer, assent to a particular dogma. These phenomena are so insistent and attention to them becomes so entrenched that it becomes difficult to break away from phenomena-based *ad hoc* categories, into the isolation of more genotypical and psychologically relevant variables.

For example, it is possible that some conversion, some church attendance, some prayer, and some belief—but not all of such instances—may share guilt-coping obsessional characteristics. Scientific progress may be gained by the development of such a category, cutting across groups of behaviors. But such progress has been slow because of the seeming dramatic insistence of the phenomena to be studied in their own terms—just as the descriptive, diagnostic categories of abnormal psychology have only slowly yielded to more “dynamic” variables, cutting across the phenomenological descriptions.

*“Religion” as a single variable.* The attempt to regard all religion as an integral phenomenon or as a single variable is the most significant and awkward illustration of the persistence of culturally convenient, but scientifically dubious, categories. A society finds it convenient and possible to group a great variety of behavior and institutions within the single category of “religion.” To its own frustration, psychological inquiry has commonly followed this practice.

Argyle (1958), for example, explicitly regards religion as a single quantifiable variable, with institutional membership or attendance, devotional practice, and orthodoxy of belief used as interchangeable indices. Other measures frequently used uncritically as a single index of “religion” are the religious-interest scale of the Allport-Vernon Scale of Values, scales such as Thurstone’s for assessing attitudes toward church, God, and Sunday observances, various tests of religious information, and self-ratings of religiosity.

Most attempts to define religion as a single variable also emphasize the uniqueness of religion as compared to other psychological and cultural phenomena. Such definitions tend to identify religion with mystical characteristics and also serve to exclude, by implication, the application to religion of psychological analyses applied to other types of behavior (Allport 1950; Clark 1958; James 1902).

But the usefulness or validity of such a global, single-variable definition of religion is not easily defended against a sophisticated theoretical viewpoint. As Johnson has dramatically argued:

In the name of religion what deed has not been done? For the sake of religion men have earnestly affirmed and contradicted almost every idea and form of con-

duct. In the long history of religion appear chastity and sacred prostitution, feasting and fasting, intoxication and prohibition, dancing and sobriety, human sacrifice and the saving of life in orphanages and hospitals, superstition and education, poverty and wealthy endowments, prayer wheels and silent worship, gods and demons, one God and many gods, attempts to escape and to reform the world. How can such diametrical oppositions all be religious? ([1944] 1959, pp. 47–48)

Several recent factor-analytic and correlational studies have suggested a single prominent factor, but it is a factor defined almost exclusively by attitudes toward religious institutions, the visible factor by which Western culture generally identifies religion. The few studies which have used more subtle indices than attitudes toward the church, and more sophisticated samples of persons more familiar with elements of religion other than institutional affiliation, have suggested that religion be considered as reflecting a multiplicity of factors (Dittes 1967).

Furthermore, empirical studies of “religion” defined as a single variable—primarily church affiliation and participation—have not tended to produce consistent, meaningful relations with other variables (perhaps because other variables are also grossly defined, in terms of such easily accessible measures as personality variables, age, education, diagnosis of pathology). Studies using such data typically show that religious involvement is “greater” among women, among young adolescents and those past middle age, among the less educated and less intelligent (when class differences are held constant), among members of the middle class, and among those scoring higher on measures of ethnocentrism and authoritarianism.

Definitions and descriptions which attempt to embrace in a single statement the religions of diverse cultures invariably seem to lose the critical and significant characteristics of any one. The number of facets of religious phenomena increases so greatly as one crosses cultural lines that cross-cultural definition and generalization are exceedingly treacherous. One perhaps ought to be pessimistic about developing a psychology of religion in general and perhaps instead learn more from studying cultural group and individual differences.

*Descriptive distinctions.* Among those committed to description of phenomena, the use of typologies has represented one attempt to define somewhat less global variables. These types are characteristically taken from those established by practitioners themselves—for example, distinctions between mystic, prophet, priest (e.g., see Clark 1958, chapters 12 and 13).



James (1902) proposed the much-quoted distinction between the religion of the healthy-minded and of the sick soul—perhaps more a distinction between personality types than between religious orientations per se. The distinction appears related to the manic-depressive continuum, with one end representing the extroverted personality ("positive thinking" in Norman Vincent Peale's terms) and the other a more depressed, introverted, openly conflicted person.

Pratt's (1920) distinction between subjective and objective worship has persisted, perhaps more as a haunting methodological dilemma than as a genuine phenomenal distinction. Objective worship, typified by the Roman Catholic mass, refers to intention and reference beyond the worshiper; subjective worship refers to self-conscious awareness of the effects of worship on the worshiper, as typified by much Protestant worship.

These typologies have not lent themselves to easy operational definitions, nor have they found themselves as part of more elaborate theoretical statements. Users of typologies have been content with typologies per se, rather than curious about the relation of the type with other variables.

Differing beliefs suggest another variable of potential psychological significance. But when these are taken in the language and the dimensions of the practitioners, they tend to emerge along a single conservative-liberal dimension, highly correlated with general "nonreligious" conservatism-liberalism and its known determinants. Other categories suggested by research in attitudes and perception outside religion, such as suggestibility, conformity, or dogmatism, may be more useful.

*Religion-religiosity.* Many writers have suggested the important distinction between "authentic" intensive inner religious experience, on the one hand, and routine peripheral religious activity, on the other. The distinction has been labeled internal-extrinsic by Adorno and others (1950), interiorized-institutionalized by Allport (1954), intrinsic-extrinsic by Allport (1959; Allport & Ross 1967), primary-secondary by Clark (1958), committed-consensual by Allen (1965; Allen & Spilka 1967), and religion-religiosity by many. The distinction reflects a value judgment and is analogous to one commonly made by religious participants themselves—especially by those regarded as prophets and reformers—who frequently call attention to the inverse relation between the two types of activity.

In empirical studies most measures used are of explicit, objective religiosity, usually involving par-

ticipation in, or attitudes toward, the formal religious institution, its overt activities, and its official doctrines. On the other hand, most theoretical efforts, including attempts at establishing distinctions between, and categories of, phenomena, are concerned with the more intrinsic, subjective religion—the inner "spiritual" life, personal attitudes and orientations, values, loyalties, and commitments. Indeed, assuming the present distinction, a major methodological weakness of most empirical research in the psychology of religion is the use of instances of extrinsic religiosity as indices of an intrinsic religious orientation (e.g., Argyle 1958).

Factor-analytic studies (e.g., Cline & Richards 1965; Allen 1965; Allen & Spilka 1967) tend to support the distinction between the explicit institution-oriented religiosity and more personal intrinsic religion. Most attempts to provide scales or other indices for each of these two elements have foundered on the likelihood that each is not a discrete variable but, like "religion" itself, a general rubric covering many independent variables. Allen (1965) and Allport and Ross (1967) report scales of extrinsic and intrinsic religion that have some construct validity (extrinsic religion was correlated with prejudice, but intrinsic religion was not).

*Psychoanalytic description.* Much psychoanalytic writing is essentially descriptive. It has been a common practice among psychoanalytical writers to identify themes thought to be of psychological significance in religious material. Religious symbols, activities, and legends have provided fertile ground in which both Freudian and Jungian adherents could discover disguised symbolic "meaning." A recent example (La Barre 1962) argues at length for the phallic symbolism of the snake in contemporary and historical cults and legends, including the Garden of Eden story. This work is classified here as largely descriptive because it focuses on the identification of a single variable rather than on the advancement of a theoretical understanding of the relation between that variable and others.

Freud's first writing on religion (1907) called attention to several descriptive similarities between religious ritual and obsessive-compulsive symptoms. The theoretical implications of common etiology and common functions have been extrapolated from this observation by others, for example, Ostow and Scharfstein (1954) and Jones ([1923] 1964, chapter 9).

*Theory-based approaches.* The two major psychological theories generally used in analyzing religious phenomena are both associated with Freud,

although neither was fully originated or developed by him. One theory emphasizes the psychological motive of dependence, the position of the mother, and the solicitous aspects of religion. The other theory emphasizes the psychological motive of hostility, the role of the father, and the demanding aspects of religion. Both theories reflect the fundamental hypothesis that the "religious life represents a dramatization on a cosmic plane of the emotions, fears, and longings which arose in the child's relation to his parents" (Jones [1923] 1964, p. 195).

The general distinction between the more solicitous and more demanding modes of religion has long been noted by many observers both within and without religion. Examples are the common adage of clergymen that their aim is to "comfort the disturbed and to disturb the comfortable" or the distinction between the creative, redemptive, justifying attributes of God—or the "saving," healing functions of religion—and the wrathful attributes of God—or the demanding, law-abiding, sanctifying functions.

Fromm (1963) has sharply contrasted these two theoretical orientations and has attempted to invoke them sequentially, in relation to changing social conditions, to account for changing emphases in early Christian centuries. Erikson (1958) has both distinguished and blended the theories in an intensive "case study" of Martin Luther's developing religious convictions. Bakan (1966) has proposed that religion represents the projection and management of *parents'* conflicts between personal "agency" motivations—including hostility to children—and nurturant motives.

*Theories based on dependence.* Dependence-based theories associated with Freud's second book on religion (1927) emphasize religion as a source of comfort, solace, or assurance in the face of external stresses and frustrations. A protective and projective parent is associated with deferred or fantasied satisfactions, especially in relation to keenly experienced frustrations. The religious response is one of submission, in imitation of such a figure as Jesus, the dutiful son.

Supporting data, beyond general reflections, are not overwhelming. Especially missing are what would be the most convincing data—correlations between evidence of significant life frustrations and submissive reliance in the framework of this particular kind of religious formulation. Some general evidence has accumulated suggesting generally "greater" religious involvement among groups such as the less intelligent and the economically deprived, who, one might infer, experience depri-

vation, frustration, and a sense of inadequacy. Evidence along these lines has been surveyed by Cronbach (1933), by Argyle (1958, pp. 145–154), and by Dittes (1967).

*Theories based on hostility.* Freud applied to religion the full range of his Oedipal theory and apparently believed that this application was confirmed by anthropological and historical materials he cited (1913; 1934–1938). Emphasis is placed on relations with a paternal god-figure and on coping with largely "inner" stresses generated by this relation, particularly hostility and rebellion. Various elements of belief and of practice are analyzed as allowing for the *expression* of such hostility, for its *control*, and for the consummating receipt of *punishment* and *forgiveness*. A prototype is the identification with a Christ-figure who assumes the role of God, yet submits to the father, is crucified, and reconciled through resurrection and ascension.

The subtle and complex variables of motivation and of religion with which this theory is concerned have so far eluded controlled empirical study.

*Projection and wish fulfillment.* Projection to God of attitudes toward a father are assumed by both theories. This assumption has given rise to empirical tests of various designs to discover whether there is a similarity between attitudes toward God and attitudes toward the actual father. The degree of correspondence discovered has varied widely, perhaps because the research has not carefully taken into consideration the theories' proposal that such projection takes place under conditions of particular personal motivation.

Wish fulfillment, or perhaps more accurately, "rationalization," is a general process postulated especially by the theories based on dependence. The degree and process by which conceptualization follows and is made consistent with motivation and practice has attracted varying formulations (e.g., Dunlap 1946). All tend to assume an independent "need to conceptualize" or a "need for consistency" among conceptualizations and between conceptualizations and behavior.

*Religion and ethnocentrism.* Perhaps the most effective fusion of empirical and theoretical work has developed from the repeated empirical finding of correlations between indices of religion and of conservative social attitudes—especially, because of the cultural importance given them in the mid-twentieth century, attitudes of anti-Semitism and racial prejudice. This correlation was brought into sharp focus and most of the subsequent developments were given direction by the work of Adorno and others (1950).



The finding has, among other things, stimulated attention to definitional problems, cited above, especially to the distinction between religiosity, which Adorno's measures reflected, and religion. Most theories, guided by an understanding of official religious doctrine, might expect religion to be directly related to liberal attitudes. [See ANTI-SEMITISM and PREJUDICE.]

Beginning with data in the Adorno volume, studies have regularly reported suggestions of a curvilinear relation between ethnocentrism and frequency of church attendance, with both non-attenders and the most faithful attenders showing more liberal attitudes than occasional attenders. If the most regular attendance can be interpreted as an index of more personal internalization of religious norms, beyond perfunctory institutional loyalty, then this may be some evidence of the theoretically expected relation between intrinsic religion and liberalism. Other evidence awaits the definition and measurement of such elements of intrinsic religion as faith, hope, love, trust, freedom, etc.

*Personality variables.* The relationship between religion and ethnocentrism has led to theory and research aimed at designating the personality characteristics that underlie both variables and mediate the relation. Research findings suggest that a person for whom both conservative ideology and conventional, institution-based religion are functional is a person with a "weak ego" or "severe superego," who is threatened by external circumstances or internal impulse and therefore likely to be responsive to the unambiguous external controls, structures, and supports provided either by a religious institution and its ideology, rituals, and moralisms or by the strict categorizing and self-enhancement provided by ethnocentric doctrines.

Such an interpretation of the function of religion is supported by a wide range of empirical reports of the correlation between religion (especially as measured by institutional loyalty and adherence to conservative doctrines) and lower intelligence, disrupted emotional life, suggestibility, constricted personality, and awareness of personal inadequacy (Dittes 1967). These findings, suggesting that religion may attract the relatively inadequate and helpless, may be less shocking than first supposed to the exponents of religious traditions whose highest religious celebrations commemorate a captivity or a crucifixion and who generally preach man's fallen state and fundamental helplessness.

An extension of this general view also seems to account for a range of dramatic phenomena, such as glossolalia, sudden emotional conversion, faith

healing, and trance states. Reports, though meager, of personality characteristics and of environmental characteristics associated with such participation suggest the existence of a general trancelike, regressive "ego constriction" that produces greater expression of impulses and greater responsiveness to suggestion and social influence. [See HYPNOSIS; HYSTERIA; PERSUASION; SUGGESTION.]

Finally, the relation between religion and ethnocentrism has stimulated further attention to defining cognitive variables and especially to making the distinction (e.g., Rokeach 1960) between the content characteristics of a belief or attitude and the content-free or structural characteristics of the style, such as the openness or the dogmatism with which it is held. This has led to the hypothesis that religion is related more to elements such as the need for closure than to the conservative elements of ethnocentrism. [See PERSONALITY, POLITICAL, article on CONSERVATISM AND RADICALISM; SYSTEMS ANALYSIS, article on PSYCHOLOGICAL SYSTEMS.]

JAMES E. DITTES

[Directly related are the entries ATTITUDES; IDEOLOGY; MYTH AND SYMBOL; RITUAL. Other relevant material may be found in ANALYTICAL PSYCHOLOGY; DEFENSE MECHANISMS; FANTASY; OBSSIVE-COMPULSIVE DISORDERS; PSYCHOANALYSIS; PSYCHOLOGY, article on EXISTENTIAL PSYCHOLOGY; and in the biographies of FREUD; JAMES; JUNG.]

#### BIBLIOGRAPHY

- ADORNO, T. W. et al. 1950 *The Authoritarian Personality*. American Jewish Committee, Social Studies Series, No. 3. New York: Harper.
- ALLEN, RUSSELL O. 1965 *Religion and Prejudice: An Attempt to Clarify the Patterns of Relationship*. Ph.D. dissertation, Univ. of Denver.
- ALLEN, RUSSELL O.; and SPILKA, BERNARD 1967 *Committed and Consensual Religion: A Specification of Religion-Prejudice Relationships*. Unpublished manuscript.
- ALLPORT, GORDON W. 1950 *The Individual and His Religion: A Psychological Interpretation*. New York: Macmillan.
- ALLPORT, GORDON W. 1954 *The Nature of Prejudice*. Cambridge, Mass.: Addison-Wesley. → An abridged paperback edition was published in 1958 by Doubleday.
- ALLPORT, GORDON W. 1959 *Religion and Prejudice*. *Crane Review* 2:1-10.
- ALLPORT, GORDON W.; and ROSS, J. MICHAEL 1967 *Personal Religious Orientation and Prejudice*. *Journal of Personality and Social Psychology* 5:432-443.
- American Journal of Religious Psychology and Education*. → Published from 1904 to 1911.
- Archiv für Religionspsychologie und Seelenführung*. → Published from 1914 to 1915, from 1918 to 1930, and from 1962 onward.
- ARGYLE, MICHAEL 1958 *Religious Behaviour*. London: Routledge. → A compendium of social-psychological

- research relating religion, conceived as a single quantifiable variable, with various social and social psychological indices.
- BAKAN, DAVID 1966 *The Duality of Human Existence: An Essay on Psychology and Religion*. Chicago: Rand McNally.
- Catholic Psychological Record. → Published since 1963.
- CLARK, WALTER H. 1958 *The Psychology of Religion*. New York: Macmillan. → A textbook primarily surveying early descriptive material.
- CLINE, VICTOR B.; and RICHARDS, JAMES M. JR. 1965 A Factor-analytic Study of Religious Belief and Behavior. *Journal of Personality and Social Psychology* 1:569-578.
- CRONBACH, ABRAHAM 1933 The Psychology of Religion. *Psychological Bulletin* 30:327-361. → A comprehensive survey of 361 publications of varying methodology.
- DITTES, JAMES E. 1967 The Social Psychology of Religion. Unpublished manuscript. → Projected for publication in the revised edition of Gardner Lindzey (editor), *Handbook of Social Psychology*.
- DUNLAP, KNIGHT 1946 *Religion: Its Function in Human Life*. New York: McGraw-Hill.
- ERIKSON, ERIK H. (1958) 1962 *Young Man Luther: A Study in Psychoanalysis and History*. Austen Riggs Monograph No. 4. New York: Norton.
- FREUD, SIGMUND (1907) 1959 Obsessive Acts and Religious Practices. Volume 2, pages 25-35 in Sigmund Freud, *Collected Papers*. International Psycho-analytic Library, No. 10. London: Hogarth; New York: Basic Books. → First published as "Zwangshandlungen und Religionsübung."
- FREUD, SIGMUND (1913) 1959 Totem and Taboo. Volume 13, pages ix-162 in *The Standard Edition of the Complete Psychological Works of Sigmund Freud*. London: Hogarth. → First published in German.
- FREUD, SIGMUND (1927) 1961 The Future of an Illusion. Volume 21, pages 5-58 in *The Standard Edition of the Complete Psychological Works of Sigmund Freud*. London: Hogarth. → First published in German.
- FREUD, SIGMUND (1934-1938) 1964 Moses and Monotheism: Three Essays. Volume 23 in *The Standard Edition of the Complete Psychological Works of Sigmund Freud*. London: Hogarth. → First published in German.
- FROMM, ERICH 1963 The Dogma of Christ. Pages 3-91 in *The Dogma of Christ, and Other Essays on Religion, Psychology, and Culture*. New York: Holt. → Summarizes the various psychoanalytic theories of religion and applies them to changing social conditions in the early Christian centuries.
- HAVENS, JOSEPH 1961 The Participant's vs. the Observer's Frame of Reference in the Psychological Study of Religion. *Journal for the Scientific Study of Religion* 1:79-87.
- HILTNER, SEWARD; and ROGERS, WILLIAM 1962 Research on Religion and Personality Dynamics. *Religious Education* 57, no. 4. → Both summarizes and evaluates a number of more recent studies; especially urges psychoanalytic and theological sophistication.
- JAMES, WILLIAM (1902) 1963 *The Varieties of Religious Experience: A Study of Human Nature*. Enlarged edition with appendices and introduction by Joseph Ratner. New Hyde Park, N.Y.: University Books. → Based largely on literary records; discusses conversion, mysticism, saintliness, and other phenomena. Establishes typologies which have endured; sets a mood and a philosophic context for the psychological study of religion.
- JOHNSON, PAUL E. (1944) 1959 *The Psychology of Religion*. Rev. ed. New York: Abingdon. → A textbook; includes a good history of the field.
- JONES, ERNEST (1923) 1964 *Essays in Applied Psychoanalysis*. Volume 2: Essays in Folklore, Anthropology and Religion. New York: International Universities Press. → Includes 8 chapters on religion summarizing compactly the fundamental psychoanalytic views of religion, plus some intriguing excursions; more compact and less polemic than Freud's own writings.
- Journal for the Scientific Study of Religion*. → Published since 1961.
- LA BARRE, WESTON 1962 *They Shall Take Up Serpents*. Minneapolis: Univ. of Minnesota Press.
- Lumen Vitae Studies in Religious Psychology*. → Published irregularly since 1959, by the Lumen Vitae Press in Brussels. These volumes contain empirical and theoretical studies which originally appeared in the journal *Lumen vitae*.
- MEISSNER, WILLIAM W. 1961 *Annotated Bibliography in Religion and Psychology*. New York: Academy of Religion and Mental Health.
- MENGES, ROBERT J.; and DITTES, J. E. 1965 *Psychological Studies of Clergymen: Abstracts of Empirical Research*. New York: Nelson.
- OSTOW, MORTIMER; and SCHARFSTEIN, BEN-AMI 1954 *The Need to Believe*. New York: International Universities Press. → An elementary but fairly comprehensive statement of Freudian theories of the social-controlling and guilt-reducing functions of ritual and other aspects of religion.
- OUTLER, ALBERT C. 1954 *Psychotherapy and the Christian Message*. New York: Harper.
- PRATT, JAMES B. 1920 *The Religious Consciousness: A Psychological Study*. New York: Macmillan.
- Review of Religious Research*. → Published since 1959.
- ROBERTS, DAVID 1950 *Psychotherapy and a Christian View of Man*. New York: Scribners. → An attempt by a theologian to integrate theological and psychological conceptions.
- ROKEACH, MILTON 1960 *The Open and Closed Mind: Investigations Into the Nature of Belief Systems and Personality Systems*. New York: Basic Books.
- STRUNK, ORLO JR. (editor) 1959 *Readings in the Psychology of Religion*. Nashville: Abingdon. → A fairly comprehensive, condensed collection; emphasizes early descriptive work.
- TILICH, PAUL 1952 *The Courage to Be*. New Haven: Yale Univ. Press. → A paperback edition was published in 1959. Illustrates an explicit attempt to incorporate psychological insights into theology.
- Zeitschrift für Religions-psychologie*. → Published from 1907 to 1913.

## RELIGIOUS OBSERVANCE

Religious observance includes all overt kinds of religious behavior. This article will consider the main sociological and psychological sources of variation in religious observance. Sociological factors include historical trends, social class, and membership of minority groups. Psychological factors include age, sex, and personality traits. We shall



consider studies carried out between 1900 and the present day in Great Britain, the United States, and western Europe. Research in this area shows the extent and patterning of religious activities in the population and makes it possible to test theories about the causes of religious behavior.

**Measures and indexes.** Frequency of church attendance is the most widely used index of religious observance. It can be assessed by actual observation, but questioning is the more usual method. If "How often do you go to church?" is asked, there is often some inflation of the true figure. "What did you do last Sunday?" is a better way of introducing the issue and gives a more accurate estimate of the behavior of a large group of people. This index is open to the further difficulties that some denominations exert greater pressure for attendance and that some people go to church for nonreligious reasons. Supplementation by other measures is therefore valuable.

One of the most important of these is the amount of private religious activity, such as saying prayers or reading the Bible. This needs to be assessed by very precisely worded questions, or again there will be exaggerations of the true amount. The profile of religious activities of different kinds may be very informative; for example, the ratio of private observance to church attendance is greater for Protestants than for Roman Catholics, and the same is true for women as compared with men. Church membership is useful for making historical comparisons, but the criteria for membership are different for different denominations, Roman Catholics and Jews being the most lenient. Furthermore, these criteria change with time: the apparent American Protestant revival in the first quarter of this century was due to changes in the definition of membership, with the result that more members were counted (National Council . . . ). Other sociological indexes include contributions to church funds (in relation to average income) and numbers of publications on religious topics, both of which can be used to study historical changes. Attitude scales, measuring attitudes toward the church or toward religion, correlate well with church attendance. Indeed, factor analyses of various indexes and measures suggest that frequency of attendance is the most representative single measure; it is also the one most often used in research.

**Variations in religious observance.** Variations in religious observance raise questions of the type, Why does A go to church while B does not? The answer may take three possible forms.

**Social learning and group pressures.** If a person is brought up in an isolated community where everyone takes part in the same religious practices, there is no difficulty in predicting his religious activities. In general a person will be greatly influenced by his family, especially if he has a close relationship with them and lives at home: correlations between attitudes of parents and children who are at student age are in the range of .40 to .60. People are also influenced by the social norms of groups to which they belong outside the home. If family and friends conflict, the influence of each is weakened, oscillation between them may be the basis for conversions and reconversions, and the final outcome will depend on which attachment is the stronger as well as on how far the orientation of each fits the individual's personality.

**Individual differences in personality.** Every ideology and its associated practices meet some needs and fail to meet others. For example, a very aggressive person is more likely to support capital punishment than pacifism. The same is true of religion, though the needs and personality mechanisms involved are many and complex (Argyle 1958; 1964). These can be deduced from findings about the religious practices and beliefs of people with different personalities.

**Socioeconomic categories.** Pressures to conformity exist within all social groups, although personality may cause some individual variations in behavior. But statistical differences in religious observance are also found to exist according to social class, region, and other socioeconomic categories. Such differences can often be accounted for by the different psychological situations of people in these groupings, for whom religious practices will have a very varied appeal. In other cases there may be a historical explanation.

### Sociological variables

**Denominational differences.** For our purposes a fourfold classification of religious observance is most useful. First, *conservative religion* includes Roman Catholics, Lutherans, and some Episcopalians and Anglicans. Members of this group, as well as having certain characteristic beliefs, have a high rate of public worship combined with a low rate of private worship. The clergy have considerable power and are held in high regard; services are formal and ritualistic. Second, the main *Protestant churches*, such as the Presbyterians and Methodists, have less ritual and formality of worship, place more emphasis on sin and guilt, and are more active in evangelism. Individual clergy-

men have their followers, but they have less formal power. The level of church attendance of Protestants is much lower than that of Roman Catholics in most countries, but when Protestants are in a minority, as in France, they actually have higher attendance rates than Roman Catholics (Fogarty 1957). Third, *sects* are at the furthest end of a dimension that has Roman Catholics at one extreme and Protestants in the middle; services are informal and highly emotional, and there is little church organization. Finally, *liberal religion* includes Unitarians, most Jews, Quakers, and some people of no religious affiliation at all. Church attendance is low in this group and is neither ritualistic nor emotional, just as the religious attitudes of this group are more cognitive than emotional.

It is impossible to provide any brief explanation of these differences. It could be said that each type of religion has arisen under specific social conditions and attracts a particular type of person. There is also a developmental relation between them. Sects represent the primitive form that emerges spontaneously under the influence of a leader; they gradually become more formal and develop into churches. Both sects and Protestant churches are splinter groups from established churches.

An example of the differences between religious bodies is provided by nationwide surveys of public and private observance in Great Britain (Table 1).

**Table 1 — Public and private religious observance in Great Britain, by denomination, in per cent**

	Weekly church attendance (1947)	Regular private prayer (1950)
Roman Catholics	52	68
Nonconformists	14	43
Church of England	8	48

Source: Adapted from Argyle 1958, p. 159.

**Historical factors.** Fairly extensive data on religious observance in the United States and Great Britain since 1900 are available from records and social surveys. In Great Britain weekly church attendance dropped from 11.5 per cent of the population to 3.5 per cent over the period from 1900 to mid-century. Membership and proportion of average income donated fell 50 per cent. This decline was, however, largely confined to the Protestant churches; the Roman Catholics showed little change. Analysis of membership and Easter communicants shows a continuous decline to 1950, followed by an increase to 1957; it has subsequently leveled off (Argyle 1958). In the United States there was a slow drop in membership (corrected to allow for

changing definition) from 1900 to 1943, and analysis of donations during this same period shows a 50 per cent drop in percentage of income donated. However, both donations and church attendances started rising in 1943 and have continued to do so. The proportion of Roman Catholics in the United States has increased from 16 per cent to 22 per cent of the population since 1900, and there has been a great expansion in a number of small sects, although many of these sects have become more like Protestant churches during this period (Thomas 1963). In France there has been a slow but steady decline in religious observance since 1850, especially in the urban areas that have become most heavily industrialized. This trend has resulted in the very great differences between town and country reported below. About 94 per cent of the population are baptized Roman Catholics, but only about 40 per cent of them attend Sunday Mass regularly (Boulard 1954).

What is the explanation of these historical changes—that is, a general decline since the second half of the nineteenth century and a temporary recovery in the 1950s? The growth of cities, industrialization, and the development of a culture that is primarily materialistic and scientific are probably some of the factors. There has been considerable variation in economic prosperity since 1900, but contrary to the expectation of some people, this appears to have no regular relationship with religious observance. There was no increase in religious observance during the depression (Kincheloe 1937); on the other hand, there has been an increase during postwar prosperity. It is true that the American sects expanded during the depression, but they also expanded before and after it. It has been alleged that the postwar increase in religious activity is due to prosperity, but the increase in Great Britain stopped in 1957, despite continuing prosperity; in any case, stronger evidence is needed to support this theory. Does war affect religious observance? There is no clear evidence of nationwide changes owing to war. On the other hand, men involved in military action often pray when in danger, and many become more religious later—though sometimes in a non-institutional manner—while others become less religious.

The decline in religious observance from 1900 to midcentury may well be due to changes in child rearing (Bronfenbrenner 1958). Protestantism has an appeal to personalities with strong superego and guilt feelings, while Catholicism appeals to people who are dependent on authority. Trends in child



rearing have made both types of personality less common. However, the increase in religious activity in the United States and Great Britain since mid-century is more difficult to explain. It is noteworthy that this increase has been in the more liberal type of religion and that other religious bodies have also moved in this direction. One factor may be the greater emphasis the churches have put on social activities, with a corresponding appeal to the emergent teen-age group, but at present there is no accepted explanation.

**Social class.** There are excellent social survey data concerning class differences. In Great Britain it is consistently found that the higher the social class, the more frequent is religious observance. According to a survey taken by Odham's Press in 1947, weekly church attendance was reported by 13.4 per cent of the working-class respondents, 16.3 per cent of lower middle-class, and 19.4 per cent of upper-class and upper middle-class respondents (cited by Gorer 1955). Similar results, though with smaller differences, are found for private observance and affiliation. In the United States there is a similar relationship between religious activity and social class: in one recent survey 34 per cent of the Protestant, semiskilled workers were regular church attenders, compared with 47 per cent of professional people; the corresponding figures for Roman Catholics were 66 per cent and 81 per cent (Lazarwitz 1961). In France rather greater differences are found in the same direction; about 10 per cent of manual workers are regular attenders at Mass as compared with about 50 per cent of managerial and professional workers (Pin 1956).

There are also class differences in denomination. In Great Britain the Church of England draws more members from higher up the social scale, while the reverse is true of the Roman Catholic and Nonconformist churches. In general, the sects appeal to the most underprivileged, although this is not true of Christian Science. The differences are not great: 55 per cent of upper middle-class people claim to be Anglican versus 48 per cent of working-class people. There are also differences between members of individual churches in a given denomination. In the United States, while over-all class differences in observance are smaller than in England, class differences in denomination are greater. There is a definite hierarchy, with Episcopalians, Presbyterians, Congregationalists, and Methodists at the top, followed by Lutherans, Baptists, and Roman Catholics, and certain sects at the bottom of the socioeconomic system. There is also a spread within each church; Roman Catho-

lics and Methodists, in particular, are drawn from all classes. In a particular town or area the class segregation may be fairly sharp: a survey of part of California found that the percentage of unskilled laborers was 84 per cent among Pentecostals, 19 per cent among members of the Assembly of God, 15 per cent among Baptists, and 1 per cent among Congregationalists (Goldschmidt 1944). In European countries, such as Germany, the Protestant minority are of higher social status than the Catholics (Fogarty 1957).

These findings can be summed up by saying that the upper and middle classes are more orthodox and formalistic, while the working classes are more emotional in their religion. Such differences can also be found within the same denominational groups (Vernon 1962).

It has been shown in the United States (Pope 1942) and observed in other parts of the world that small sects usually start among the very poor. As the members become more prosperous, the church changes its character to become more like an established Protestant church. The less prosperous members then break away to form another sect, which goes through the same cycle again. Social mobility of a whole group of people may thus result in a change in the character of their church. Mobility of an individual, however, especially upwards, may lead to his changing to another denomination.

The single most important fact here is that poor people are not, in general, more religious than others. This clearly contradicts the many theories, of which those of Marx and Freud are the best known, according to which religious belief is the outcome of social frustration. However, such theories might help to explain the sect type of religion, in which tension-discharging meetings and beliefs with a strong element of compensatory fantasy are typically accompanied by the expectation of an imminent end of the world and consequent improvement in the lot of the members. The explanation of the changes in churches as their members move up the social scale is simply that as people become more prosperous and settled their need for tension-reducing services and compensatory fantasies declines. Thus it may be that the Church of England has consistently alienated the industrial working classes of Great Britain by its middle-class and generally conservative outlook; this, at any rate, would partly account for its relatively small working-class membership.

The higher social class of the members of the major Protestant churches may be explained by

Max Weber's theory that Protestantism develops certain attitudes conducive to capitalism, perhaps through Protestant child rearing, which engenders greater achievement motivation.

**Minority groups.** Minority groups are often underprivileged and show an above-average level of religious activity, especially in sect religion, as in the instance of the American Negroes. African Negroes also have a taste for emotionally expressive types of religion. Another sociological factor here may be that of being uprooted from the rural or tribal community: new arrivals at the outskirts of large towns are found to be active in sect religions in Europe and Canada as well as in Africa (Mann 1955). Finally, immigrant minority groups may import their previous religious practices, which are retained and encouraged in order to bind the group together: indeed, the church may act as a kind of community center.

Turning to otherworldly matters on the part of underprivileged minority groups has been criticized on the grounds that it does not help to solve their real grievances and indeed distracts attention from them (Myrdal 1944). In some cases, however, churches and their leaders have taken part in non-violent resistance and other political activities designed to advance the interests of minority groups. A further explanatory principle that should be mentioned in this connection is the theory of the quest for ego identity. Erikson (1956) has postulated a need for a clear and satisfactory picture of the self; and this picture is partly based on the reactions of others. Thus uprooted migrants, who have no place in society, can be greatly helped by sect membership, where up to 15 per cent of the members can be officeholders (Mann 1955).

**Town and country.** Surveys from several countries show a clear difference between town and country: the larger the community, the lower the level of religious observance. This tendency is strong enough to counteract another factor in the United States—that is, the tendency of Roman Catholics to be concentrated in the cities. In Great Britain the weekly attenders vary from 12 per cent in large cities to 17 per cent in small towns (Gorer 1955). In France there is a rather greater difference between town and country; in country areas about 60 per cent are regular attenders of Mass, compared with 25 per cent or less in the cities (Boulard 1954; Pin 1956). The most probable explanation is simply cultural lag: the slow decline in religion in the first half of the century has not yet affected rural areas. A further probability is that the churches are still maintaining a rural

world picture and have not yet come to terms with the modern industrial world; this is particularly true of the Roman Catholic church in France.

**The mass media and evangelism.** Public meetings are the traditional means of evangelism. They have not wholly been replaced by radio and television and are not likely to be. Their great advantage over less direct means of communication is their ability to arouse emotional excitement; there is considerable psychological evidence to suggest that for the change of deep-rooted attitudes and beliefs, emotional arousal is very effective—especially when combined with a persuasive message and social pressure (Sargant 1957). Studies of evangelical meetings show that the higher the emotional temperature, the greater the percentage of people converted. At Billy Graham's campaigns in Great Britain, 5.3 per cent made decisions at the larger meetings, compared with 2.3 per cent at smaller ones and with 0.8 per cent at overflow meetings. However, when people are converted by highly emotional methods, the proportion of those who are permanently converted is less. About one-half of Graham's converts (excluding those who were church attenders already) were still active a year later. The comparable figure for the nineteenth-century revivals was one in six or less (Starbuck 1899).

The conditions for conversion at a public meeting are as follows: large, crowded meetings; a highly prestigious preacher, skilled at arousing emotional states; a message showing how a change of behavior or belief will reduce the anxiety or guilt aroused; and pressure exerted by assistant evangelists. Those most likely to be converted are females between 13 and 21 years of age who have strong guilt feelings and no background of religious instruction.

The explanation of religious conversion takes two forms, depending on whether it is "gradual" or "sudden." Gradual conversions are nowadays the more common and can be interpreted in terms of social influence and social learning: for example, exposure to the religious ideas of an educational institution or peer group. Sudden conversions usually occur during adolescence and usually among people with strong guilt feelings. What probably happens is that young people's guilt feelings are heightened by the preacher; they are then especially receptive to Protestant doctrines of salvation that promise relief of guilt feelings.

It is not known to what extent religious behavior and beliefs can be affected by radio or television. The experience with political and other attitudes



is that actual conversion is rare, though there can be minor changes and the creation of beliefs on new topics, while exposure over a long period can bring about a slow erosion of attitudes that previously flourished unchallenged.

### Psychological variables

**Sex differences.** Women are more active in religion than men, but the extent of the difference varies with the level of activity, the denomination, and the type of observance. In the United States, where there is a high level of observance, about 50 per cent of all women and 42 per cent of all men attend some place of worship every week, giving an attendance ratio of 1.19 women for every man. In Great Britain there is a lower level; the corresponding percentages are about 17 for women and 11 for men, giving a ratio of 1.55 women per man. One possible explanation of this difference in ratios is that when there is more observance, conformity pressures are greater, so that individual differences are reduced.

The sex ratio also varies with denomination. In the United States the ratio of female to male members varies from about 1.10 for Roman Catholics and Lutherans to 1.35–1.55 for the main Protestant denominations to 1.55–2.10 for the sects. This can be explained partly in terms of the conformity pressures mentioned above: when the clergy are more powerful, the rate of attendance is greater and individual differences less.

The sex ratio varies with types of observance. The ratio is least for the public aspects of worship, greatest for the more private. In the United States the ratio of women to men for saying private prayers is about 1.65, compared with 1.19 for church attendance. Again, this is what would be predicted by the conformity theory.

It has not yet been explained why there should be a differential sex ratio at all. There are two possible lines of explanation. First, women have stronger consciences and stronger guilt feelings and therefore respond more strongly to Protestantism. One study finds a strong correlation between guilt and frequency of church attendance among Protestant women (Argyle 1964). Second, when God is presented as like a father, he should have a greater appeal for women, if the Oedipus theory is true. The smaller sex ratio for Roman Catholics would be expected, since they are provided with a mother figure as well as a father figure. Thus both of these theories may account for the greater sex ratio among Protestants.

**Age differences.** Religious observances and beliefs are different at each stage of life. In child-

hood (3–10) there is considerable activity, including praying for the gratification of immediate needs, holding fairy-tale religious beliefs, and sharing the religious activities of the family. In adolescence (10–18) there is often a good deal of conflict, as a result of exposure to differing influences and ideas. It is not so much a period of religious activity as a period of decision. Although most conversions to religion occur during this period, it is probable that there is an equal number of reconversions, since there is no over-all increase in religious observance. Religious observance at this age, however, is marked by rather greater emotional fervor than at other ages, and young people will readily take part in the conversion of others.

During young adulthood (18–30) there is a slow decline in religious observance: 25–35 per cent of those religiously active at 18 are inactive at 30. On the other hand, a few people have their first conversion during this period. During later adulthood (30 onward), there is a steady increase in religious activity, so that virtually 100 per cent of people of 90 and over are religiously active (Cavan et al. 1949). This increase has not been found in some recent American surveys, probably because the increase in observance since 1943 has masked aging factors as deduced from cross-sectional surveys. However, the increase with age has been shown by surveys at other times and places, as has the definite reversal in trend at 30–35. Among older people both the emotional and intellectual aspects of religion are feeble, and observances have long become habitual. They often take an active part in the organizational aspects of church life [see AGING, article on SOCIAL ASPECTS].

The explanation of the religion of later adulthood is the easiest: older people become anxious about death, and religion helps to reduce this anxiety. There is extensive support for this view: old people who are religious are better adjusted, and as people become older they increasingly give reassurance of immortality as their reason for church attendance. The observance of adolescents is differently motivated: by the relief of guilt mechanism; by the projection of the superego onto God or onto the clergy; and by cognitive needs to understand basic problems of life, for which only religion provides answers. The relatively low rate of religious observance of the 30–35 age group is due to the relative absence of these motivations, combined with a preoccupation with earthly problems involving occupation, family, and money. The decline in observance among Roman Catholics in the 30–40 age group has been attributed to a further factor—conflicts about birth control.

**Personality factors.** Although the findings on the subject of personality factors as related to religious activity are rather scattered and inconsistent, certain generalizations can be put forward. First, Protestants may be more guilt-ridden and intrapunitive than Roman Catholics. This is supported by the finding that Protestants have a higher suicide rate than Catholics (Dublin & Bunzel 1933); although this difference is related to the fact that suicide is forbidden by the Roman Catholic church, this fact itself needs to be explained. Moreover, Protestants have lower crime rates, especially for aggressive crimes, as compared with Roman Catholics; their mental patients are more often depressed (Farr & Howe 1932) and their converts are preoccupied with guilt before the event (Starbuck 1899). Second, in general, churchgoers are more conformist, suggestible, authoritarian, and right-wing in politics than others. This is particularly true of the members of the larger established churches; it is least true of Quakers, members of other churches of the liberal type, and sect members (Adorno et al. 1950). Third, young religious people are somewhat more neurotic than others, while old religious people are rather better adjusted. Finally, among psychotic patients, about one in seven is preoccupied with religious ideas or practices (Farr & Howe 1932), and many religious leaders and innovators have been found to have psychotic symptoms.

My own investigations of religious activity indicate that two psychological mechanisms may account for a number of these findings. The first of these two mechanisms has been introduced already: the relief of guilt feelings by the acceptance of Protestant doctrines of salvation. Frequency of church attendance has been found to correlate with strength of guilt feelings for women only (Argyle 1964). It seems likely that this mechanism operates for female Protestants, especially between the ages of 14 and 30, and that this accounts for the findings on neurosis and suicide.

The second theory states that internal conflict between the self and the superego results in the latter becoming projected or externalized onto clergy, saints, or God. In one study Roman Catholics were found to have a greater conflict between self and superego than others, and it has been shown that they make greater use of projection as a defense mechanism (Dreger 1952). This may account for the greater power of the Roman Catholic clergy, the relatively greater amount of Roman Catholic public observance, and the more severe Roman Catholic image of God.

MICHAEL ARGYLE

[See also RELIGION; SECTS AND CULTS.]

#### BIBLIOGRAPHY

- ADORNO, THEODOR W. et al. 1950 *The Authoritarian Personality*. American Jewish Committee, Social Studies Series, No. 3. New York: Harper.
- ARGYLE, MICHAEL 1958 *Religious Behaviour*. London: Routledge.
- ARGYLE, MICHAEL 1964 Seven Psychological Roots of Religion. *Theology* 67, no. 530:333-339.
- BOULARD, FERNAND (1954) 1960 *An Introduction to Religious Sociology: Pioneer Work in France*. London: Longmans. → First published as *Premiers itinéraires en sociologie religieuse*.
- BRONFENBRENNER, URIE 1958 Socialization and Social Class Through Time and Space. Pages 400-425 in *Society for the Psychological Study of Social Issues, Readings in Social Psychology*. 3d ed. New York: Holt.
- CAVAN, RUTH S. et al. 1949 *Personal Adjustment in Old Age*. Chicago: Science Research Associates.
- DREGER, RALPH M. 1952 Some Personality Correlates of Religious Attitudes as Determined by Projection Techniques. *Psychological Monographs* 66, no. 3.
- DUBLIN, LOUIS I.; and BUNZEL, BESSIE 1933 *To Be or Not to Be: A Study of Suicide*. New York: Smith & Haas.
- ERIKSON, ERIK H. 1956 The Problem of Ego Identity. *Journal of the American Psychoanalytic Association* 4:58-121.
- FARR, C. B.; and HOWE, R. L. 1932 The Influence of Religious Ideas on the Etiology, Symptomatology, and Prognosis of the Psychoses: With Special Reference to Social Factors. *American Journal of Psychiatry* 11:845-865.
- FOGARTY, MICHAEL P. 1957 *Christian Democracy in Western Europe: 1820-1953*. Univ. of Notre Dame (Ind.) Press.
- FREUD, SIGMUND (1930) 1958 *Civilization and Its Discontents*. Garden City, N.Y.: Doubleday. → First published as *Das Unbehagen in der Kultur*.
- GOLDSCHMIDT, W. R. 1944 Class Denominations in Rural California Churches. *American Journal of Sociology* 49:348-355.
- GORER, GEOFFREY 1955 *Exploring English Character*. New York: Criterion.
- KINCHELOE, S. C. 1937 *Research Memorandum on Religion in the Depression*. Social Science Research Council, Bulletin No. 23. New York: The Council.
- LAZARWITZ, B. 1961 Some Factors Associated With Variations in Church Attendance. *Social Forces* 39:301-309.
- LOOK MAGAZINE 1955 *A Guide to the Religions of America: The Famous Look Magazine Series on Religion, Plus Facts, Figures, Tables, Charts, Articles, and Comprehensive Reference Material on Churches and Religious Groups in the United States*. Edited by Leo Rosten. New York: Simon & Schuster.
- MANN, WILLIAM E. 1955 *Sect, Cult and Church in Alberta*. Univ. of Toronto Press.
- MYRDAL, GUNNAR (1944) 1962 *An American Dilemma: The Negro Problem and Modern Democracy*. New York: Harper. → A paperback edition was published in 1964 by McGraw-Hill.
- NATIONAL COUNCIL OF THE CHURCHES OF CHRIST IN THE UNITED STATES OF AMERICA *Yearbook of American Churches*. → Published since 1916, under various titles. The National Council is the successor to the



Federal Council of the Churches of Christ in America. See especially the 1956 issue.

- PIN, ÉMILE 1956 *Pratique religieuse et classes sociales dans une paroisse urbaine, Saint-Pothin à Lyon*. Paris: Éditions Spes.
- POPE, LISTON 1942 *Millhands and Preachers: A Study of Gastonia*. New Haven: Yale Univ. Press; Oxford Univ. Press.
- SARGANT, WILLIAM W. (1957) 1959 *Battle for the Mind: A Physiology of Conversion and Brain-washing*. Rev. ed. London: Pan Books. → A new edition was published in 1961 by Penguin Books.
- STARBUCK, EDWIN D. (1899) 1911 *The Psychology of Religion: An Empirical Study of the Growth of Religious Consciousness*. 3d ed. New York: Scribner.
- THOMAS, JOHN L. 1963 *Religion and the American People*. Westminster, Md.: Newman.
- VERNON, GLENN M. 1962 *Sociology of Religion*. New York: McGraw-Hill.

## RELIGIOUS ORGANIZATION

Religious organization is the complex of institutionalized roles and procedures which regulate the relations of men with the supernatural order, however such an order may be conceived. It may involve the regulation of religious practice and the promulgation of true, and suppression of false, doctrine; procedures for the recruitment, education, and professional socialization of religious specialists; the structure of authority among them; the basis, extent, and nature of their authority over the laity and its territorial and temporal coordination; the control of religious places, periods, premises, and equipment and such property and temporalities as accrue in the establishment of this control.

Despite the wealth of religious scholarship, there are very few studies of religious organization, and of this exiguous literature very little employs sociological analysis. Organization theory as developed in sociology has little relevance to non-Western religions. In Western Christianity and Judaism, religious roles and procedures constitute articulated structures which, with increasing self-consciousness, approximate to bureaucratic models; however, even in the West, the extent of this approximation is limited by the inherent nature of the religious quest. The patterns of affective neutrality, role specificity, performance expectations, self-orientation, and even universalism, which characterize the dominant organizational mode of Western society, are all, in greater or lesser degree, alien to religious institutions, roles, relationships, and values. The comparative historical analysis of Max Weber, although not specifically concerned with organization as such, and the more typological stud-

ies of Joachim Wach remain the outstanding works in this field.

In non-Western societies, religious activities are usually an intimate part of the life of the local community and lack an independent and articulated organizational structure. There may be loose association between communal activities and those of regional religious centers. Local religious agents operate within the community, and their responsibilities do not extend beyond it. Family, kinship, tribal, local, and even national cults represent a high degree of convergence of political and religious institutions; such cults were found widely in ancient civilizations, as documented by William Robertson Smith (1889), N. D. Fustel de Coulanges (1864), C. W. Westrup (1939-1954), Marcel Granet (1922), Martin P. Nilsson (1941-1950), and (more sociologically) by Max Weber (1922a) and Joachim Wach (1944). The development of a national cult is apparent in the Old Testament, and Bellah (1957, chapter 4) makes evident its importance in Japanese Shintoism.

**Religious authority in East and West.** Religious roles are universal, but they are not always performed by full-time specialists. The part-time functionary has less opportunity than the full-time specialist to acquire authority and thus can be less adequately integrated into an organization. Autonomous religious authority is likely to arise only where the mystical and magical offices of religious functionaries are regarded by secular authorities as indispensable for the legitimation of their power. This was a function of the Brahmins in early Hinduism, but they did not evolve a religious organization of the rational hierarchic type; the preeminence of religious status may itself have limited such a development. The Brahmin caste was a hereditary guild which monopolized religious functions; religious power tended to be localized and particular, diffused within the wider society but not subjected to division of labor and hierarchization within the stratum.

In contrast, Christian priests lacked this ascriptive basis of status; the hereditary principle was checked by the imposition of celibacy; charisma was attached to the office rather than to the person; and appointment to office was often dependent on political agencies. Under these circumstances, religious functionaries often stood in a relationship of uneasy interdependence with political authorities. Religious authority, limited in its exercise over secular society, developed as a hierarchic structure within the religious institution itself. The religious institution gained coherence by modeling itself on the political structure, which, in certain circum-

stances, it came to supersede or transcend. The articulation of responsibility within the religious institution necessarily involved insistence on the coherence and exclusiveness of doctrine and liturgical practice. This insistence created divisions in Christianity after political and coercive power was lost to the church, but the principles of coherence and exclusiveness became characteristic of all the separate Christian organizations.

### Leaders, members, and organization

An important variable affecting religious organization is the extent to which religion is founded on prophecy. Certain religions, such as Christianity, rely on conversion to the teachings, ritual, ethics, and example of the prophets who founded them. In religions which do not trace their foundation to a prophet, allegiance is determined largely by ethnicity, though even within the traditions of founded religions, insulated religious communities arise in which birthright and, hence, ethnicity become the principal, if not exclusive, criteria of eligibility for participation (Hostetler 1963; Francis 1948). The degree of voluntary commitment which they demand clearly differentiates founded from unfounded religions. Christianity almost always requires voluntary commitment, however formal this may be in practice. In traditional Hinduism, by contrast, voluntary acceptance of ritual obligations by the unentitled would be inconceivable, while their neglect by Hindus, with the implied rejection of caste obligations and the status claims of Brahmins, could bring only spiritual perdition and social ostracism (Weber [1921] 1958, pp. 8 ff.; Hutton [1946] 1964, chapters 6 and 7). Although the convergence of ethnic, social, and religious status in Hinduism was challenged by the development of distinctive cult practices, such as the Lingayat (Weber [1921] 1958, pp. 298 ff.), these cults also lacked systematic organization and have frequently been vulnerable to reabsorption into the caste structure.

**The case of Judaism.** In Judaism after the Diaspora, the destruction of the Temple and the disappearance of priestly authority made religious practice the distinguishing criterion of Jewish identity and the focus of ethnic and community values. Jewish communities were organized around the synagogue, which became the surrogate of the Temple, as did the rabbi of the priest. Rabbinical patriarchs superseded the priestly hierarchy, but once they were deprived of political power and dispersed in relatively small communities, the principal function of rabbis became that of interpreters of the law, and it was from this function that they

derived their authority as religious and community leaders. A high degree of uniformity of faith and order rested on principles drawn from the synagogical service. Whereas the prophet legitimated his authority in terms of divine inspiration and the priest in terms of hereditary right to perform sacrifice, the rabbi's claim rested on his knowledge of the law, which covered ritual, taboos, and civil matters, such as marriage, divorce, and inheritance.

Thus, the rabbi's role traditionally had community significance rather than specifically religious significance. This fact and the fact that the rabbi's role is open to all Jewish males who undergo appropriate instruction are conditions which limited the development of hierarchy, especially when we take into account the dispersal of Jews into small communities. At a relatively late date, rabbis in dispersed European communities formed wider associations, but even today "there is . . . little formal hierarchical organization . . . and the role itself remains functionally diffuse . . ." (Carlin & Mendlovitz [1958] 1960, p. 379). Discrimination and persecution enhanced the significance of synagogical organization and of religious ritual and law for European Jewry, whereas in less intolerant contexts, Jews let their observance of law and ritual decline and became more fully assimilated to the host society (Strizower 1962). In industrial societies, modern Judaism has increasingly adopted the type of organization characteristic of Christian denominationalism (Steinberg 1965). This process has been facilitated by the fact that authority resides very largely in the individual congregation (Sklare 1955, pp. 40-42). Rabbinical authority has declined as the rabbi's functions have grown less relevant to modern life. Although Judaism is today essentially a nonconversionist religion, the secularization of Jewry has led to extensive diversification of religious commitment on the part of individual Jews and given modern Judaism some of the characteristics more typical of founded religions.

**The problem of authority.** All religions have had the problem of enforcing their own practices and eliminating indigenous magical preoccupations. The acceptance of the ethical prescriptions of Buddhism originally implied the cenobitic life, and initially Buddhism lacked organizational precepts for the control of the laity; thus its development as a lay religion depended on political support. Without a theology or a liturgy, Buddhism, in its lay expression, was less protected than most religions from the accretion of indigenous magical traditions; hagiolatrous, idolatrous, and logolatrous tendencies developed, especially in Mahāyāna Buddhism. Abbots acquired religious and political con-



trol of lay populations, most notably in the lama-series of Tibet.

Christianity, with its strong authority structure, and in the context of an increasingly rational climate of thought, was widely successful in eliminating magical practices and in enforcing religious observance—sometimes, where it has dominated political affairs, by punitive measures. But even Christianity has been less successful in its mission territories; and among the country people of even long-settled Christian areas, and among unacculturated urban groups, magical ideas persist, especially in non-Protestant countries (Siegel 1941; Bastide 1960).

### Hierarchy and the priestly function

Religious institutions may attain a high degree of organizational, doctrinal, and ritualistic consistency from the association of religious roles with political or legal roles or in imitation of political structures.

**Christian hierarchies.** The Christian church, growing up outside the political framework of the Roman Empire but later becoming affiliated to it, evolved an autonomous hierarchy. Its organization withstood the subsequent fragmentation of the previously supporting political structure and in some circumstances assumed political authority. In contrast to the profuse variation evident in Far Eastern religions at local and regional levels, Christianity, in spite of its numerous divisions, appears inherently more monolithic in structure. For example, simultaneous allegiance to Taoism, Confucianism, and Buddhism has occurred in China and, with the addition of Shinto, in Japan. Among the world's religions, organization based on a universalistic principle of rigorously exclusive voluntary belief is a characteristic only of Christianity and Islam (Judaism, being ethnically based, is not universalistic).

Christianity inherited exclusiveness from Judaism but soon accepted the universalistic principle of voluntary commitment. Early Christians were organized as self-chosen autonomous congregations of saints, probably under colleges of presbyters. Priestly functions, partially derived from Hebrew sources, developed with presbyters and deacons instead of the Aaronic priests and Levites, according to the *Epistle to the Church at Corinth* of Pope Clement I (fl. A.D. 96). Whether the early leaders were congregationally chosen or were itinerant apostolic deputies, and what rights congregations had over them are questions which are disputed. Some local leaders, such as Ignatius of Antioch (fl. A.D. 115), claimed considerable eminence and from metropolitan centers acquired

authority over surrounding areas. Thus a monar-chic type of episcopacy developed. Bishops were only gradually subordinated to the authority of metropolitan centers and finally to the papacy (Telfer 1962). The hierarchy of the imperial state influenced the development of religious hierarchy. Struggles between rival centers, which often espoused mutually heretical doctrines, finally resulted in increased centralization and coordination, the growth of which was also facilitated by the metaphysical doctrine of one universal church, of which particular congregations were seen as localized expressions. Hierarchy thus extended both above and below the episcopal level, and the pope came to appoint men to bishoprics (unless a temporal authority usurped this power) and instituted the metropolitan title *legatus natus*, which was normally conferred on archbishops as a reinforcement of papal control.

The modern Roman church declares that a bishop has sovereignty in his diocese; that the episcopate is of divine right; that the pope cannot arrogate episcopal rights; and that bishops have not become his instruments. However, it also asserts that the pope, as the universal ordinary, has power over all churches and that bishops owe personal allegiance to the holy see (Rahner 1964; Rahner & Ratzinger 1961; Telfer 1962). The nicety of theological argument cannot conceal the reality of papal power; indeed, the laity as such has no influence in the appointment of bishops in the Roman church. However, in the autocephalous Greek Orthodox church the laity retains vestiges of at least nominal control, although real power is in the hands of the bishops, whose executive functions are vested in an annually appointed synod of 12 of their number under the permanent presidency of the archbishop of Athens (Hammond 1956, chapter 4). In the Roman, Orthodox, and Anglican churches, the religious hierarchy stands above the laity and gives these churches the character of objective, external, and transcendent institutions (Troeltsch [1912] 1931, pp. 331–343; Yinger [1957] 1961, chapter 6).

**Islam—a society under law.** Although Islamic society was theocratic, the absence of priestly roles meant that hierarchization occurred in its political rather than in its religious institutions, which never acquired autonomy from the political system established in the name of the prophet. Political power was consolidated with religious sanction, toward the achievement of a new integration of society transcending the tribal level (Watt 1961; Levy [1931–1933] 1957, chapter 1). The prophet's successors, the caliphs (*imāms*), were, in orthodox

Sunni theory, exponents of the divinely inspired law, and their function was to preserve religion and exercise leadership (Ibn Khaldūn [1375–1382] 1958, vol. 1). In early Islam caliphs held the office of judge, later delegated, and came to exercise royal authority at certain periods. The Sunni caliphs enjoyed no sacredness of character; they were never popes, as medieval Christians supposed (Daniel [1960] 1962, p. 225) and as was supposed of the sultan of the Ottoman Empire in the eighteenth century. (Some Shi'ite sects regard the caliphate as a divine institution, but their attempts to secure power for the descendants of Ali met with little success: the Fatimids of Egypt were the only important Shi'a caliphs.) Hierarchization occurred in Islam in the administration of the sacred law (*shari'ah*) and not in connection with specifically religious functions.

Islam has always been a community under the law as much as a society with a particular religion. Orthodoxy was determined not by councils but by the consensus (*ijmā'*) of the learned doctors (*'ulamā'*) who administered the law within the four interpretative traditions which developed. In the Ottoman Empire, political and religious institutions were one in theory and closely related in fact. Religious functionaries administered education and law, but no autonomous religious organization arose. Distinction between various functionaries never became rigid: the *'ulamā'* were teachers, preachers, legists, and magistrates (*qādī*), sometimes interchanging roles. As payment and pensions were increasingly granted to *'ulamā'* in principal cities, they passed increasingly under centralized executive control. Spiritual affairs, conceived in terms of the sacred law, eventually passed to the supreme jurisdiction of the *mufti* (jurisconsult) of Istanbul, who, as officially appointed head of the leading school of Islamic law, provided legitimation for the sultan's authority (Gibb & Bowen 1950–1957, chapter 9).

But even this development did not provide the basis for an independent religious authority structure. The hereditary principle persisted at the fringe of Islamic social structure, in the ranks of the *Shārif* (supposed descendants of the prophet), and even in the local office of *mufti* (an office unlike that of *qādī* in that it relied on fees for legal consultation and not on a stipend for the discharge of official duties). But since Islam lacked specific priestly functions, the hereditary principle did not bring into being a sacerdotal guild. In all religions, the sacramental element of priestly office appears to be a necessary condition for the development of an independent and centralized religious hierarchy.

In Islam the main factors preventing the development of such a hierarchy appear to have been the restriction which developed on the interpretation of the law; the minimal liturgical character of religious practice and the absence of sacerdotalism; the unrestricted eligibility of all men to perform religious functions; and the establishment of social control through hierarchically organized agents whose functions emphasized the legal rather than the sacred aspects of religiously given law.

### Religious administration and polity

If a founded religion is to survive, it must evolve an organization that will outlast its charismatic founder. Max Weber termed this process "the routinization of charisma." New roles of teacher and legist evolve and are institutionalized. The new functionaries elaborate and codify the prophet's teachings, and priests, if priestly office exists, superintend the development of rituals and liturgies.

**Monks and laymen.** These very general characteristics of founded religions should be borne in mind when we compare East and West. Buddhism, for instance, is vested in a religious order, the members of which neither superintend parishes nor serve congregations. Theravāda Buddhism in its pure form makes few concessions to the laity: the true disciple of the Buddha is not the householder, but the monk. Salvation is remote, and complete renunciation of the world is an ideal practically impossible of attainment. Monks perform ceremonies for laymen, and act as objects of that charity by which laymen can accumulate merit; but the meditation in which, ideally, the monk is engaged is a more meritorious activity. "... The Buddhist clergy are not in charge of the laity. . . . Having retired from the world for self-improvement, the monk expects the layman to come to him" (Wells [1939] 1960, p. 11). The rituals that monks undertake for laymen in Thailand, Ceylon, Burma, and Japan are neither intercessions nor sacramental acts. The hierarchies that have evolved, for example, in Thailand, where Buddhism has been closely associated with the monarchy, are essentially internal to the monastic order itself, even though hierarchs perform state ceremonies (for example, in Thailand), exert political pressure (in Ceylon and Vietnam), or even, as in certain periods in Japan, exercise some political control. In the Theravāda countries, lay Buddhist associations have developed, modeled on Western types of organization and, in Ceylon, directly due to the influence of the Theosophists. In Tibet, where monachistic theocracy developed, and where Tan-



tric magic deeply penetrated Buddhist practice, a religious hierarchy evolved which established lamaism as a distinctive type of Buddhist theocracy. The lamas represented multiple, vicarious, and constantly renewed manifestations of saviorhood in their persons, not by any ordination from above, but by virtue of birth and discovery. Ecclesiastical organization was centered around these manifestations of the Buddhahood, and lamas assumed regulation of the political order. Regulations found in religious writings were intended primarily for monks, and although extended in some measure to the laity, there was no formal code prescribing religious observances for the Tibetan. Monks and lamas did not supervise morals, but offered solace, contact with clairvoyant wisdom, and thaumaturgical power (Ekvall 1964).

**The Christian parish.** In Christianity, bishops divided their dioceses and delegated their sacerdotal authority; thus local units were established, in each of which a priest served the population. Parish organization was the comprehensive territorial provision of religious facilities and religious control. Parochial cures derived from episcopal divine right, and parish priests were supposed to serve all residents in their district, all of whom had religious obligations. The theoretical assumptions implicit in this ecclesiastical organization depended ultimately on the principle of political support for religious monopoly. Parish organization was transformed once the secular authorities permitted religious diversity and even permitted bishops of different churches to claim spiritual jurisdiction over the same territory. Once this occurred, religious movements became voluntary associations on what was effectively a denominational pattern, even though the traditional churches retained the styles and ecclesiastical theories of the older pattern of organization.

**Congregationalism.** Congregational polity rests on different bases of authority—democratic rather than authoritarian and hierarchic. The religious leader is called by the congregation of the self-chosen. He may be accredited by a seminary or by a college of other ministers, but there is always an assumption of spiritual equality among the believers and political equality in the management of their affairs. Furthermore, the laity may assume sufficient knowledge and divine guidance to resist false teaching. Authority beyond the local level may be ill-defined, even in advanced societies with good systems of communication, although local congregations of those who accept a particular theological position are almost everywhere in some form of loose association. Where religious and political institutions are closely associated (as they were in

Islam for much of its history and in seventeenth-century New England Congregationalism), political officials may acquire general superintendence of religious faith and practice.

A necessary condition for congregational autonomy is the absence of sacerdotalism. The religious leader is an administrator, perhaps also a preacher and a pastor. He enjoys neither monopoly nor charisma of office. He may, as in Judaism, acquire enhanced social status by virtue of scholarship or piety, and these may be the demands of the role. All religious roles are achieved rather than ascribed, and there is open recruitment. Where strong differences of education exist between minister and congregation and are reinforced by the establishment of such institutions as seminaries, law schools, and the Islamic divinity school (*madrasah*), tendencies for ministerial and lay expressions of faith to diverge may be noted. If the religion of the learned becomes too intellectual or formal, popular movements, often reasserting the validity of experiential religion, arise in congregational polities as readily as in church organizations (for example, Hassidism in eighteenth-century Judaism and Sufism in Islam). Alternatively, congregational control often diminishes in practice, and an intellectual leadership may introduce new doctrines and liturgies—a process illustrated in the rise of Unitarianism among Presbyterian and Congregationalist ministers in the late eighteenth and early nineteenth centuries.

**Synodical government.** Presbyterianism is a semiconservational polity in which the synodical authority of ministerial and lay representatives at regional and denominational levels is combined with safeguards to preserve some autonomy for local congregations. In the Church of Scotland a complex structure of checks and balances operates to give maximum democratic representation (Highet 1960). A variant pattern of synodical government was forged, though not without schisms, in English Methodism, but with the balance of authority persisting in the denominational conference. Other modern denominations have adopted the same type of system.

**Congregationalism and sacred law.** Congregational polity is legitimated by appeal to Holy Scripture as an objective body of law superior to the authority of institutions—such as the Roman church—which claim a monopoly on scriptural interpretation. Congregational groups usually accept a brief creed stating their understanding of the central teachings and requirements of scripture. This is typical of the Protestant demand for “the priesthood of all believers” (that is, a nonsacramental priesthood)—a demand which relies on

open access to the Scriptures, widespread education, individual right of comment within a broad interpretative framework, and a democratic polity of the kind established by the Congregationalists and Baptists. However, both of these denominations have a professional ministry, in continuity with historic Christian tradition, and this has permitted division of labor and ministerial education, which has grown increasingly specialized. In the absence of a clear authority structure, the professional minister in congregationalist denominations is subject to acute role conflicts and contradictions of status. Clergy role problems are not of course confined to congregational polities, though they appear to be most acute there. Perhaps because this type of ministry is prevalent in the United States these problems have been the subject of numerous sociological investigations (Fichter 1961, chapter 6; Gustafson 1954; Blizzard 1958).

The most radical congregational groups reject a professional ministry, since they believe that any man may minister to others. In the Methodist church, lay ministry became widely established, though it arose from secular democratic pressures rather than in response to a specific theological orientation. It had been developed earlier among the Quakers; it developed as "mutual ministry" among the Disciples; and in some sects, notably the Plymouth Brethren and the Christadelphians, it has been accompanied by a persistent antiprofessionalism.

**Islamic organization.** Islam has always manifested the equality of believers, and any sane male adult may lead others—ideally a quorum of 40—in prayer. In Ottoman Islam, various classes of mosque officials arose in the big mosques; the *imām*, who technically deputized for the caliph, was an *'alim* (learned doctor). With him was often associated a preacher and also a *khatīb*, who delivered two harangues on Fridays; a *mūazzin* to call the faithful to prayer was an indispensable, but usually unpaid, official. The *qādī*, mistaken for a bishop by many Christian writers, administered the *sharī'ah*, and the *muhtasib* served as public censor, ensuring among other duties the maintenance of public morality and the upkeep of the mosque (Levy [1931–1933] 1957, pp. 330 ff.; Gibb & Bowen 1950–1957, pp. 95 ff.). In rural areas today, although the mosque is losing many of its extrareligious functions, it is still the center for ceremonies and charitable activities. Such a mosque usually belongs to the community, but it is the regular attenders who elect elders to look after it. The local *imām* is the spokesman of the *'ulamā'*, but there are no formal channels of communication, and "the various *'ulamā'* neither appoint nor discharge village

*imāms*. This responsibility rests finally with the members of the mosque" (Barclay 1964, p. 165). At village level, the *imām* may get no pay and may have other means of making a living, in much the same way as a parish priest in Greek Orthodoxy. In big city mosques, however, where government or endowment supports religious activities, there is usually a permanent *imām*.

### Religious control of property

Religious organization, even in millenarian sects, involves ownership of property. In the case of Hinduism, the Brahmins initially acted as chaplain-magicians to princes but came increasingly to own, or at least to control, the temples and shrines (Weber 1921). A prebendary system evolved, and members of the priestly class came to acquire property of their own, to which were extended the attributes of sanctity and inalienability. Since Brahmins were a religious stratum and were not hierarchically organized, such property became the property of families or local organizations.

In Western Christianity, by contrast, property came to be owned by a unified church corporation once the Hildebrandine reforms had established a celibate clergy. Indeed, the Roman church, in spite of disendowment in various countries and periods, is one of the largest corporate property owners in the world, even if one leaves out of account its specifically undeployable wealth in the form of church premises and religioartistic assets. This wealth is controlled by specialized agencies which the hierarchy has empowered, and their functions are not open to lay scrutiny. The power of these agencies and the extent to which economic action is dictated by spiritual or political considerations have never been subjected to sociological analysis. The church is a large shareholder in many, especially Italian, enterprises and operates in the capital markets of the world. The invested wealth of the Roman church is thought to be more than \$5,000 million.

The Church of England has assets of about \$560 million; it has entrusted its financial concerns to a body of commissioners, who operate on the stock market in much the same fashion as other investors. The profits from their operation are used partly for subsidization of clerical stipends, which are a more significant issue for the largely non-celibate Anglican clergy than in the Roman church, from whose celibate clergy more complete allegiance can be commanded, and upon whom it can impress more effectively the virtue of frugality. All of the financial operations of the Church of England are open to parliamentary inquiry.

The absence of hierarchy in other religions, and



of a rational financial structure in non-Western societies, has restricted the accumulation of corporate deployable wealth as it has occurred in Christianity, although local religious functionaries have frequently controlled large proportions of local property. Religious premises are often locally controlled, either, as is partially the case in the Church of England, under vested rights of local patrons or through local religious authorities. In the Church of England the right to present priests to "livings" has tended (if we exclude crown and government patronage) to pass increasingly from private to episcopal control (Mayfield 1958; Paul 1964, pp. 195-197). In Islam, local communities are often responsible for mosques, and the local *imām* may persuade regular attenders to undertake repair work on the mosque. The endowment system—a consequence of the emphasis on almsgiving—became extensive in Islam and of such proportions that some governments have maintained special ministries to supervise the *waqf* (endowments) which support mosques, especially in the cities (Guillaume [1954] 1961, pp. 178 ff.; Gibb & Bowen 1950-1957, chapter 12; Smith 1957, p. 205).

Protestantism has usually left control of property to local trustees, but denominational control and the centralization of financial administration have been extended. This appears to have gone furthest in connexional bodies, such as the American Methodist church (Muelder 1963, pp. 148 ff.).

### Church, sect, and denomination

The designation "church" defines an autonomous corporate institution hierarchically organized and served by a professional priesthood. Churches that conform to all these specifications are found only in Christianity; the structural arrangements of other religions are best referred to as religious institutions.

The concept "sect," however, has wider application. In a religion which lacks central organization, members who feel themselves to be religiously deprived and who therefore seek new facilities, new activities, and re-estimations of their social worth can usually be accommodated by the ready accretion of beliefs and practices to the dominant tradition. Only where the social and religious status of established functionaries is threatened are such demands likely to be universally resisted; this may be the occasion for the emergence of a sect—that is, a new religious movement which rejects the authority of the dominant religious tradition.

Although sects arise not only from distinct processes of schism, we may refer to a sect whenever the relations between the new movement and the

one from which it has seceded are characterized by mutual rejection and exclusion (reserving the term cult for movements in which such exclusiveness is absent). This process occurs most typically where the dominant religion is centrally organized. But even in centralized religions, rejection of sectarian elements may occur piecemeal—witness the gradual rejection of Methodism by the Church of England or Methodism's own steady development of sectarian features toward the end of Wesley's life and in the decades following his death. "Sect," then, is a concept of wider applicability than "church," even though the sharpest contrast occurs in the case of Christianity (Troeltsch 1912; Brewer 1952; Wilson 1963).

**Denominationalism.** The sect exists in a relationship of tension with the wider society. Over time this tension may diminish, and as this occurs the sect may evolve into a denomination (Niebuhr 1929). This is by no means an inevitable process; it is true only of certain types of sects and is more conspicuous in the United States than in Europe (Wilson 1959). It has also been maintained (Martin 1962) that not all denominations have evolved from sects. Denominationalism is the typical form of religious organization in the pluralist industrial society. The denomination does not claim universal allegiance coterminous with the state, as does the church; it claims less institutional objectivity. Like a sect, it emphasizes voluntary adherence, but it claims no monopoly of truth and demands much less of its adherents in the way of doctrine or morals. As an institutional form it evidences the diminution and compartmentalization of religious influence in industrial society. Even where it inherits the hierarchic authority structure of the church, it offers more lay opportunity in government. Its religious leadership, in contrast with that of most sects, tends to grow more professional; as Paul M. Harrison (1959) has shown, denominations have a tendency to become bureaucratized.

**Cults.** Denominationalism implies a degree of exclusiveness and inner coherence usually lacking in non-Christian religions. Hinduism gives the individual a choice of cults that emphasize a particular style of devotion to one or another deity (Eliot 1921, chapter 25; Wach 1944, pp. 127-130). Christianity has also evolved devotional cults to particular saints within the Roman church and even personal followings for occasional charismatic individuals, but such movements are more fully coordinated with more regular practices and are subjected to established procedures of discipline and superordinate central authority. Centrally organized, stable movements of a nonexclusive type,

which are, however, more distinctive than church parties, have also arisen in Protestantism: Moral Re-Armament, Anglo-Israelism, and the Faith Mission are examples (Eister 1950; Wilson 1967).

**Monasticism and lay orders.** Monasticism, in which high religious commitment was emphasized by the renunciation of property, sexuality, and personal independence, was the very basis of early Buddhism, and it has been of great importance in Christianity, both under the direction of organized orders in the Roman church and in a more diffuse form in the Orthodox churches. However, it is of reduced significance today. Monasticism, the *collegium pietatis*, and the inner fraternities (for which Wach 1944, pp. 174–186 provides a typology) exist only with the support of the mass of the laity. The maintenance of that support has become an increasing problem for all religions and has tended to direct them from monastic to other patterns of religious commitment. This concern for lay support was, according to Weber ([1921] 1958, pp. 238–240), a principal factor in the transformation of Buddhism from a monastic to a missionary religion.

Among Muslims the development of lay orders (*ṭarīqah*) has demonstrated the need for more immediate allegiances and a stronger emotional orientation than the dominantly legalistic and unritualistic monotheism of the faith provides. Evans-Pritchard (1949, chapters 1–3) sees the Sanusiya in Cyrenaica as a reintegrating agency in an otherwise divisive society. Although the *ṭarīqah* emphasize their orthodoxy, they establish new structural features and superimpose new units of polity, sometimes maintaining their own schools, rituals, and meeting houses. They are not exclusive; members of different *ṭarīqah* may, according to H. B. Barclay (1964, p. 170), pray together in the same mosque, and some men may belong to more than one *ṭarīqah*. Through the lodges (*zawīyah*) new local hierarchies are created, which wield effective power often on the hereditary principle; around this hereditary lineage a cenobitic fraternity arises. The order uses religious feeling as the basis for more dynamic forms of social control. The 'ulamā' themselves became incorporated in the more orthodox orders and this alliance of orthodoxy with Sufism "supplied [it with] a concrete organization which spread over all ranks of society and found a place in it for every member" (Gibb & Bowen 1950–1957, vol. 1, part II, p. 78). The Muslim orders provide an emotional outlet similar to that afforded by the cults of saints which are a widespread accretion to Islam in north Africa (Gellner 1963).

## Religious training and recruitment

The recruitment and training of religious specialists is a field in which very little sociological study has been undertaken. The ascriptive basis of priesthood in traditional Hinduism and Judaism sustained hereditary guilds which monopolized religious functions. Training was thus an apprenticeship in ritual, the mysteries of which were protected by the sanctified social status of its practitioners. In Hinduism the novice was personally subordinated to his teacher's authority. The diversification of Hindu cults and philosophies meant that the postulant for religious experience (usually of a private, ecstatic nature) need no longer be a Brahman but could learn from the *guru* of his choice. Buddhist monks maintain a novitiate with open access and egress; classic Buddhism expects no more of the monastic novice than that he should learn the *dharma*, and accept ten precepts which call for a somewhat more exacting renunciation of worldly attachments than that expected of the laity. The ordained monk rehearses the 227 rules of monastic life. The Buddhist monk of Ceylon and other Theravāda countries is seeking his own salvation, although he does set an example for others and provides them with opportunities of acquiring merit (Ames 1964). He is not a priest, takes no vows, administers no sacraments, and is not prevented from leaving the order at will: "he has no other sanctity than attaches to his own good living" (Coomaraswamy [1916] 1928, pp. 154–155). In Theravāda Buddhism, particularly as it is practiced in Burma and Thailand, it has been usual for boys to spend some time in a monastery as part of their regular education (Eliot [1921] 1954, vol. 3, p. 69; Wells 1939; Nash 1965).

In Christianity, clerical education has evolved along increasingly rational lines, but a sense of vocation (or "calling") is regarded as crucial to occupational selection and training. The Roman church not only stresses the sense of vocation but also presents training for the priesthood as a gradual induction into mysteries which culminates in the acquisition of priestly charisma. Throughout Christianity priesthood is conceived as a lifelong commitment, and ordination is a permanent investment of sacerdotal authority. At the episcopal level, the Roman, Orthodox, Old Catholic, and Anglican churches emphasize apostolic succession. Scholarly education has, however, steadily grown as a requisite of clerical office in Christianity, and there has been, especially in Protestantism, a tendency to reduce the emphasis on the exclusively mystical aspects of clerical func-



tions. As universities have been secularized, clerical education has tended to shift to theological institutions. These are conducted, even within Protestantism, with an intensification of moral rigor that often approaches the monastic, thus achieving a more intense socialization to the clerical role (Fichter 1961). In the Greek Orthodox church only a small percentage of the parish priests have more than a few years of any sort of education, though recently a number of Orthodox seminaries have been established.

### Bureaucratization in the modern church

Administrative centralization has been a conspicuous process in contemporary Christianity. As movements have sought to take advantage of modern techniques and procedures for promoting their mission, they have been obliged to adopt bureaucratic and professionalized organization. The inappropriateness of bureaucracy for the diffuse, affective, and collectivity-oriented values of faith has not prevented its adoption by church organizations (see Etzioni 1961, part I; Paul 1964; Wilson 1965).

The tensions of the situation have been experienced most acutely by those denominations with strong theological dispositions to congregational polity and lay control. Congregationalism itself was basically anticlerical rather than specifically anti-professional, but the shift of decision making from local to denominational level and the development of departmental specialists with far-reaching powers have given rise to fear of a return of an episcopacy. Theological considerations frequently sanctify particular organizational forms or proscribe certain institutional arrangements. This is most evident in the case of sects (Wilson 1967); however, the same dogmatic support underlies church organization, and, though less uncompromisingly, that of denominationalism (see, for instance, Harrison 1959). The analysis of the relationship of ideology and institutional arrangements still awaits the attention of sociologists of religious organization.

BRYAN R. WILSON

[Directly related are the entries RELIGION; RELIGIOUS OBSERVANCE; RELIGIOUS SPECIALISTS; SECTS AND CULTS. Other relevant material may be found in the articles on BUDDHISM; CHRISTIANITY; HINDUISM; ISLAM; JUDAISM; SOCIAL MOVEMENTS; and in the biographies of TROELTSCH; WEBER, MAX.]

### BIBLIOGRAPHY

- AMES, MICHAEL M. 1964 Magical-animism and Buddhism: A Structural Analysis of the Sinhalese Religious System. *Journal of Asian Studies* 23 (Supplement): 21-52.
- BARCLAY, HAROLD B. 1964 *Buuri al Lamaab: A Suburban Village in the Sudan*. Ithaca, N.Y.: Cornell Univ. Press.
- BASTIDE, ROGER 1960 *Les religions africaines au Brésil: Vers une sociologie des interpénétrations de civilisations*. Paris: Presses Universitaires de France.
- BELLAH, ROBERT N. 1957 *Tokugawa Religion: The Values of Pre-industrial Japan*. Glencoe, Ill.: Free Press.
- BLIZZARD, SAMUEL W. 1958 The Protestant Parish Minister's Integrating Roles. *Religious Education* 53: 374-380.
- BREWER, EARL D. C. 1952 Sect and Church in Methodism. *Social Forces* 30: 400-408.
- CARLIN, JEROME E.; and MENDLOVITZ, SAUL H. (1958) 1960 The American Rabbi: A Religious Specialist Responds to Loss of Authority. Pages 377-414 in Marshall Sklare (editor), *The Jews: Social Patterns of an American Group*. Glencoe, Ill.: Free Press.
- COOMARASWAMY, ANANDA (1916) 1928 *Buddha and the Gospel of Buddhism*. London: Harrap; New York: Putnam.
- DANIEL, NORMAN (1960) 1962 *Islam and the West: The Making of an Image*. Edinburgh Univ. Press.
- EISTER, ALLAN W. 1950 *Drawing Room Conversion: A Sociological Account of the Oxford Group Movement*. Sociological Series, No. 6. Durham, N.C.: Duke Univ. Press.
- EKVALL, ROBERT B. 1964 *Religious Observances in Tibet: Patterns and Functions*. Univ. of Chicago Press.
- ELIOT, CHARLES NORTON (1921) 1954 *Hinduism and Buddhism: An Historical Sketch*. 3 vols. London: Routledge; New York: Barnes & Noble.
- ETZIONI, AMITAI 1961 *A Comparative Analysis of Complex Organizations: On Power, Involvement, and Their Correlates*. New York: Free Press.
- EVANS-PRITCHARD, E. E. (1949) 1954 *The Sanusi of Cyrenaica*. Oxford: Clarendon.
- FICHTER, JOSEPH H. 1961 *Religion as an Occupation: A Study in the Sociology of Professions*. Univ. of Notre Dame Press.
- FRANCIS, EMERICK K. 1948 The Russian Mennonites: From Religious to Ethnic Group. *American Journal of Sociology* 54: 101-107.
- FUSTEL DE COULANGES, NUMA DENIS (1864) 1956 *The Ancient City: A Study on the Religion, Laws, and Institutions of Greece and Rome*. Garden City, N.Y.: Doubleday. → First published in French.
- GELLNER, ERNEST 1963 Sanctity, Puritanism, Secularization and Nationalism in North Africa: A Case Study. *Archives de sociologie des religions* 15: 71-86.
- GIBB, HAMILTON A. R.; and BOWEN, HAROLD 1950-1957 *Islamic Society and the West: A Study of the Impact of Western Civilization on Moslem Culture in the Near East*. 2 parts. New York: Oxford Univ. Press.
- GRANET, MARCEL (1922) 1951 *La religion des Chinois*. 2d ed. Paris: Presses Universitaires de France.
- GUILLAUME, ALFRED (1954) 1961 *Islam*. 2d ed., rev. Baltimore: Penguin.
- GUSTAFSON, J. M. 1954 An Analysis of the Problem of the Role of the Minister. *Journal of Religion* 34: 187-191.
- GUSTAFSSON, BERNDT 1963 Types of Religious Institutionalization. Pages 123-143 in World Council of Churches, Commission on Faith and Order, Study Commission on Institutionalism, *Institutionalism and Church Unity*. Edited by Nils Ehrenstrom and Walter G. Muelder. London: Student Christian Movement; New York: Association Press.
- HAMMOND, PETER 1956 *The Waters of Marah: The Present State of the Greek Church*. London: Rockliff.

- HARRISON, PAUL M. 1959 *Authority and Power in the Free Church Tradition: A Social Case Study of the American Baptist Convention*. Princeton Univ. Press.
- HIGHET, JOHN 1960 *The Scottish Churches: A Review of Their State 400 Years After the Reformation*. London: Skeffington.
- HOSSETLER, JOHN A. 1963 *Amish Society*. Baltimore: Johns Hopkins Press.
- HUTTON, JOHN H. (1946) 1964 *Caste in India: Its Nature, Function and Origins*. 4th ed. Oxford Univ. Press.
- IBN KHALDUN (1375-1382) 1958 *The Muqaddimah: An Introduction to History*. 3 vols. New York: Pantheon. → First published in Arabic. Contains a selected bibliography.
- LEVY, REUBEN (1931-1933) 1957 *The Social Structure of Islam*. 2d ed. Cambridge Univ. Press.
- MARTIN, D. A. 1962 *The Denomination*. *British Journal of Sociology* 13:1-14.
- MAYFIELD, GUY (1958) 1963 *The Church of England: Its Members and Its Business*. 2d ed. Oxford Univ. Press.
- MOBERG, DAVID O. 1962 *The Church as a Social Institution: The Sociology of American Religion*. Englewood Cliffs, N.J.: Prentice-Hall.
- MUELDER, WALTER G. 1963 *Problems of Church Bureaucracy*. Pages 145-168 in *World Council of Churches, Commission on Faith and Order, Study Commission on Institutionalism, Institutionalism and Church Unity*. Edited by Nils Ehrenstrom and Walter G. Muelder. London: Student Christian Movement; New York: Association Press.
- NASH, MANNING 1965 *The Golden Road to Modernity: Village Life in Contemporary Burma*. New York: Wiley.
- NIEBUHR, H. RICHARD (1929) 1954 *The Social Sources of Denominationalism*. Hamden, Conn.: Shoe String Press. → A paperback edition was published in 1957 by World.
- NILSSON, MARTIN P. 1941-1950 *Geschichte der griechischen Religion*. 2 vols. Munich: Beck.
- PAUL, LESLIE 1964 *The Deployment and Payment of the Clergy: A Report*. London: Church Information Office.
- RAHNER, KARL 1964 *Bishops: Their Status and Functions*. London: Burns & Oates. → First published as *Über den Episkopat*.
- RAHNER, KARL; and RATZINGER, JOSEPH (1961) 1962 *The Episcopate and the Primacy*. New York: Herder. → First published in German.
- SIEGEL, MORRIS 1941 *Religion in Western Guatemala: A Product of Acculturation*. *American Anthropologist* New Series 43:62-76.
- SKLARE, MARSHALL 1955 *Conservative Judaism: An American Religious Movement*. Glencoe, Ill.: Free Press.
- SMITH, WILFRED CANTWELL 1957 *Islam in Modern History*. Princeton Univ. Press.
- SMITH, WILLIAM ROBERTSON (1889) 1956 *The Religion of the Semites: The Fundamental Institutions*. New York: Meridian. → First published as the first series of *Lectures on the Religion of the Semites*.
- STEINBERG, STEPHEN 1965 *Reform Judaism: The Origin and Evolution of a "Church Movement"*. *Journal for the Scientific Study of Religion* 5:117-129.
- STRIZOWER, SCHIFRA 1962 *Exotic Jewish Communities*. London and New York: Yoseloff.
- TELFER, WILLIAM 1962 *The Office of a Bishop*. London: Darton.
- TROELTSCH, ERNST (1912) 1931 *The Social Teaching of the Christian Churches*. 2 vols. New York: Macmillan. → First published as *Die Soziallehren der christlichen Kirchen und Gruppen*.
- WACH, JOACHIM (1944) 1951 *Sociology of Religion*. Univ. of Chicago Press.
- WARNER, WELLMAN J. 1930 *The Wesleyan Movement in the Industrial Revolution*. London: Longmans.
- WATT, W. MONTGOMERY 1961 *Islam and the Integration of Society*. London: Routledge.
- WEBER, MAX (1921) 1958 *The Religion of India: The Sociology of Hinduism and Buddhism*. Glencoe, Ill.: Free Press. → First published as *Hinduismus und Buddhismus*, Volume 2 of *Weber's Gesammelte Aufsätze zur Religionssoziologie*.
- WEBER, MAX (1922a) 1963 *The Sociology of Religion*. Translated by Ephraim Fischer. Boston: Beacon. → First published as "Religionssoziologie" in Part 2 of *Weber's Wirtschaft und Gesellschaft*.
- WEBER, MAX (1922b) 1952 *Ancient Judaism*. Glencoe, Ill.: Free Press. → First published as *Das antike Judentum*, Volume 3 of *Weber's Gesammelte Aufsätze zur Religionssoziologie*.
- WEBER, MAX (1922c) 1951 *The Religion of China: Confucianism and Taoism*. Glencoe, Ill.: Free Press. → First published as "Konfuzianismus und Taoismus," Volume 1 of *Weber's Gesammelte Aufsätze zur Religionssoziologie*.
- WELLS, KENNETH E. (1939) 1960 *Thai Buddhism: Its Rites and Activities*. Bangkok: Privately published.
- WESTRUP, C. W. 1939-1954 *Introduction to Early Roman Law: Comparative Sociological Studies*. 5 vols. Copenhagen: Munksgaard; Oxford Univ. Press.
- WILSON, BRYAN R. 1959 *An Analysis of Sect Development*. *American Sociological Review* 24:3-15.
- WILSON, BRYAN R. 1961 *Sects and Society: A Sociological Study of the Elim Tabernacle, Christian Science, and Christadelphians*. Berkeley: Univ. of California Press.
- WILSON, BRYAN R. 1963 *Typologie des sectes dans une perspective dynamique et comparative*. *Archives de sociologie des religions* 16:49-63.
- WILSON, BRYAN R. 1965 *The Paul Report Examined*. *Theology* 68:89-103.
- WILSON, BRYAN R. (editor) 1967 *Patterns of English Sectarianism*. London: Heinemann.
- YANG, CH'ING-K'UN 1961 *Religion in Chinese Society: A Study of Contemporary Social Functions of Religion and Some of Their Historical Factors*. Berkeley and Los Angeles: Univ. of California Press.
- YINGER, J. MILTON (1957) 1961 *Religion, Society and the Individual: An Introduction to the Sociology of Religion*. New York: Macmillan.

## RELIGIOUS SPECIALISTS

- I. ANTHROPOLOGICAL STUDY  
II. SOCIOLOGICAL STUDY

Victor W. Turner  
Ivan A. Vallier

I  
ANTHROPOLOGICAL STUDY

A religious specialist is one who devotes himself to a particular branch of religion or, viewed organizationally, of a religious system. "Religion" is a multivocal term whose range of meanings varies in different social and historical contexts. Never-



theless, most definitions of religion refer to the recognition of a transhuman controlling power that may be either personal or impersonal. A religious specialist has a culturally defined status relevant to this recognition. In societies or contexts where such power is regarded as impersonal, anthropologists customarily describe it as *magic*, and those who manipulate the power are magicians. Wherever power is personalized, as deity, gods, spirits, daemons, genii, ancestral shades, ghosts, or the like, anthropologists speak of *religion*. In reality, religious systems contain both magical and religious beliefs and procedures: in many of them the impersonal transhuman (or mystical, or nonempirical, or supernatural) power is considered to be a devolution of personal power, as in the case of the mystical efficacy or rites established *in illo tempore* by a deity or divinized ancestor.

**Priest and prophet.** Scholars have tended to distinguish between two polarities of religious specialization. Max Weber, for example, although well aware of numerous historical instances of their overlap and interpenetration, contrasts the roles of priest and prophet. He begins by making a preliminary distinction between priest and magician. A priest, he writes, is always associated with "the functioning of a regularly organized and permanent enterprise concerned with influencing the gods—in contrast with the individual and occasional efforts of magicians." Accordingly, the crucial feature of priesthood is that it represents the "specialization of a particular group of persons in the continuous operation of a cultic enterprise, permanently associated with particular norms, places and times, and related to specific social groups." In Weber's view, the prophet is distinguished from the priest by "personal call." The priest's claim to religious authority derives from his service in a sacred tradition; the authority of the prophet is founded on revelation and personal "charisma." This latter term has been variously defined by Weber (in some contexts it seems almost to represent the *Führerprinzip*), but it may broadly be held to designate extraordinary powers. These include, according to Weber, "the capacity to achieve the ecstatic states which are viewed, in accordance with primitive experience, as the preconditions for producing certain effects in meteorology, healing, divination and telepathy." But charisma may be either ascribed or achieved. It may be an inherent faculty ("primary charisma") or it may be "produced artificially in an object or person through some extraordinary means." Charisma may thus be "merited" by fastings, austerities, or other ordeals. Even in such cases, Weber asserts,

there must be some dormant capacity in the persons or objects, some "germ" of extraordinary power, already vested in them. The prophet, then, is a "purely individual bearer of charisma," rather than the representative of a sacred tradition. He produces discontinuity in that cultic enterprise which it is the priest's major role to keep "in continuous operation." Weber's prophet feels that he has a "mission" by virtue of which he "proclaims a religious doctrine or divine commandment." Weber refuses to distinguish sharply between a "renewer of religion" who preaches "an older revelation, actual or suppositious" and a "founder of religion" who claims to bring completely new "deliverances," for, he says, "the two types merge into one another." In Weber's view, the charisma of a prophet appears to contain, in addition to ecstatic and visionary components, a rational component, for he proclaims "a systematic and distinctively religious ethic based upon a consistent and stable doctrine which purports to be a revelation" [(1922); see also CHARISMA and WEBER, MAX].

Weber's distinction between priest and prophet has its main relevance in an analytical frame of reference constructed to consider the relationship between religion as "a force for dynamic social change" and religion as "a reinforcement of the stability of societies" (Parsons 1963). It has been found effective by such anthropologists as Evans-Pritchard (1949; 1956) and Worsley (1957a; 1957b) who are dealing directly with social transitions and "the prophetic break," or what Parsons calls "the primary decision point [between] a direction which makes for a source of evolutionary change in the . . . established or traditional order, and a direction which tends either to reinforce the established order or at least not to change it drastically" (1963, p. xxix in 1964 edition).

**Priest and shaman.** Anthropologists who are less concerned than Weber with the genesis of religions and with internal developments in complex societies or their impact on the "primitive" world are inclined to contrast priest not with prophet but with shaman or spirit medium and to examine the relationship between these statuses as part of the normal working of the religious system in the simpler societies. In their excellently representative *Reader in Comparative Religion* (1958), the editors W. A. Lessa and E. Z. Vogt devote a whole section to this distinction.

Often, where there is a priest the shaman is absent, and vice versa, although both these roles may be found in the same religion, as among the Plains Indians. According to Lowie (1954), a

Plains Indian shaman is a ritual practitioner whose status is acquired through a personal communication from a supernatural being, whereas a priest does not necessarily have a face-to-face relationship with the spirit world but must have competence in conducting ritual. Lessa and Vogt ([1958] 1965, p. 410) expand these differences: a shaman's powers come by "divine stroke," a priest's power is inherited or is derived from the body of codified and standardized ritual knowledge that he learns from older priests and later transmits to successors. They find that shamanism tends to predominate in food-gathering cultures, where the shaman most frequently performs a curing rite for the benefit of one or more patients and within the context of an extended family group. Shamanistic rites are "noncalendrical," or contingent upon occasions of mishap and illness. The priest and priestly cult organization are characteristically found in the more structurally elaborated food-producing—usually agricultural—societies, where the more common ceremonial is a public rite performed for the benefit of a whole village or community. Such rites are often calendrical, or performed at critical points in the ecological cycle.

**Shaman and medium.** Raymond Firth (1964a, p. 638) regards shamanism as itself "that particular form of spirit mediumship in which a specialist (the *shaman*) normally himself a medium, is deemed to exercise developed techniques of control over spirits, sometimes including mastery of spirits believed to be possessing another medium." This definition, like that of Howells (1948), stresses the control exercised over spirits. Howells describes the shaman as "bullyragging" gods or spirits and emphasizes his intellectual qualities as a leader. This element of mastery makes the shaman a distinctive type of spirit medium, one who is believed to be "possessed by a spirit (or closely controlled by a spirit) [and who] can serve as a means of communication between other human beings and the spirit world" (Firth 1964b, p. 689). The spirit medium per se need not exert mastery; he is rather the vessel or vehicle of the transhuman entity.

Thus, although we sometimes find the two functions of priest and shaman combined in the same individual (Piddington 1950), mediums, shamans, and prophets clearly constitute subtypes of a single type of religious functionary. The priest communicates with transhuman entities through ritual that involves cultural objects and activities. The medium, shaman, and prophet communicate in a person-to-person manner: they are in what Buber (1936) would describe as an I-thou relationship with the deities or spirits. The priest, on the other

hand, is in what may be called an I-it relationship with the transhuman. Between the priest and the deity intervenes the institution. Priests may therefore be classified as institutional functionaries in the religious domain, while medium, shaman, and prophet may be regarded as subtypes of inspirational functionaries. This distinction is reflected in characteristically different modes of operation. The priest presides over a rite; the shaman or medium conducts a seance. Symbolic forms associated with these occasions differ correlatively: the symbols of a rite are sensorily perceptible to a congregation and have permanence in that they are culturally transmissible, while those of a seance are mostly in the mind of the entranced functionary as elements of his visions or fantasies and are often generated by and limited to the unique occasion. The inspirational functionary may describe what he has clairvoyantly perceived (or "been shown" as he might put it), but the institutional functionary manipulates symbolic objects with prescribed gestures in full view of this congregation.

**Sociocultural correlates.** Since the priest is an actor in a culturally "scripted" drama, it is but rarely that priests become innovators, or "dramatists." If they do assume this role it is mainly as legislative reformers—by altering the details of liturgical procedure—that they do so. If a priest becomes a radical innovator in religion, he is likely to become a prophet to his followers and a heretic to his former superiors. From the priestly viewpoint it is the office, role, and script that are sacred and "charismatic" and not the incumbent of priestly office. The priest is concerned with the conservation and maintenance of a deposit of beliefs and practices handed down as a sacred trust from the founders of the social or religious system. Since its symbols at the semantic level tend to condense the critical values, norms, and principles of the total cultural system into a few sensorily perceptible representations, the sanctification of these symbols is tantamount to a preservative of the entire culture. What the priest is and does keeps cultural change and individual deviation within narrow limits. But the energy and time of the inspirational functionary is less bound up with the maintenance of the total cultural system. His practice has more of an *ad hoc* flavor; he is more sensitive and responsive than the priest to the private and personal, to the mutable and idiosyncratic. This type of functionary thrives in loosely structured food-gathering cultures, where he deals individually with specific occasions of trouble, or during periods of social turbulence and change, when societal consensus about values is sharply declining and nu-



merically significant classes of persons and social groups are becoming alienated from the orthodox social order. The shaman subtype is completely a part of the cultural system of the food-gatherers; the prophet may well stand outside the cultural system during such a period of decomposition and propose new doctrines, ethics, and even economic values.

The shaman is not a radical or a reformer, since the society he services is traditionally flexible and mobile; the prophet is an innovator and reformer, for he confronts a tightly structured order that is moribund and points the way to religious forms that will either provide an intensified cognitive dynamic for sociocultural change or codify the new moral, ideational, and social structures that have been inarticularly developing.

There are of course significant differences in the scale of the societies in which shaman and prophet operate. The shaman enacts his roles in small-scale, multifunctional communities whose religious life incorporates beliefs in a multitude of deities, daemons, nature spirits, or ancestral shades—societies that Durkheim might have described as possessing mechanical solidarity, low moral density, and segmental organization. The prophet tends to come into his own when the division of labor is critically replacing "mechanical" by "organic" solidarity, when class antagonisms are sharpened, or when small-scale societies are decisively invaded by the powerful personnel, ideas, techniques, and cultural apparatus (including military skills and armaments) of large-scale societies. The shaman deals in a personal and specific way with spirits and lesser deities; the prophet enters into dialogue, on behalf of his whole community, with the Supreme Being or with the major deities of a traditional pantheon, whose tutelary scope embraces large numbers of persons and groups, transcending and transecting their traditional divisions and animosities. Alternatively he communicates with the generalized ancestors or *genii loci*, conceived to be a single anonymous and homogeneous collectivity rather than a structure of known and named shades, each representing a specific segment of society. Whereas the shaman's function is associated with looseness of structure in small-scale societies, the prophet's is linked with loosening of structure in large-scale societies or with incompatibilities of scale in culture-contact situations.

**Divination and religious specialists.** In its strict etymological sense the term "divination" denotes inquiry about future events or matters, hidden or obscure, directed to a deity who, it is believed, will reply through significant tokens. It usually refers

to the process of obtaining knowledge of secret or future things by mechanical means or manipulative techniques—a process which may or may not include invoking the aid of nonempirical (transhuman) persons or powers but does not include the empirical methods of science.

In the analysis of preliterate societies divination often is concerned with the immediate problems and interests of individuals and subgroups and but seldom with the destinies of tribes and nations. It is this specificity and narrowness of reference that primarily distinguishes divination from prophecy. Nadel (1954, p. 64) has called the kind of guidance it offers "mechanical and of a case-to-case kind." The diviner "can discover and disentangle some of the hidden influences which are at work always and everywhere . . . he cannot uncover any more embracing design. . . . Yet within the limits set to it divination has a part to play, providing some of the certainty and guidance required for provident action." Thus, although its range and scope are more circumscribed than those of prophecy, divination is believed to reveal what is hidden and in many cases to forecast events, auspicious and inauspicious.

Divination further refers to the analysis of past events, especially untoward events; this analysis often includes the detection and ascription of guilt with regard to their perpetrators, real or alleged. Where such untoward events are attributed to sorcerers and witches the diviner has great freedom of judgment in detecting and determining guilt. Diviners are frequently consulted by victims' relatives and show intuitive and deductive virtuosity in discovering quarrels and grudges in their clients' kin groups and local communities. Social anthropologists find important clues to areas and sources of social strain and to the character and strength of supportive social norms and values in the diviners' diagnoses.

There is evidence that mediums, shamans, and priests in various cultures have practiced divination. The medium and shaman often divine without mechanical means but with the assistance of a tutelary spirit. In the work of Lessa and Vogt there is a translation of a vivid first-person account by a Zulu informant of a diviner's seance. This mediumistic female diviner

dramatically utilizes some standard procedures of her art—ventriloquism, prior knowledge of the clients, the overhearing of the client's unguarded conversation, and shrewd common sense—to enable her spirits to provide the clients with advice. In this example, . . . a boy is suffering from a convulsive ailment. The spirits discover that an ancestral spirit is spitefully causing

the boy's illness: the spirits decree that the location of the family's village must be moved; a goat must be sacrificed to the ancestor and the goat's bile poured over the boy; the boy must drink *Itongo* medicine. The treatment thus ranges from physical to social actions—from propitiation of wrathful ancestors to prescription of a medicinal potion. (Lessa & Vogt [1958] 1965, p. 340)

Similar accounts of shamanistic divinatory seances have been recorded by anthropologists working among North and South American Indians, Eskimos, and Siberian tribes, in many parts of Africa, and among Afro-Americans.

Divination was a function of members of the priesthood in many of the complex religious systems of Polynesia, west Africa, and ancient Mexico; in the religions of Israel, Greece, Etruria, and Rome; in Babylonia, India, China, Japan, and among the Celts. According to Wach,

The Etruscans made these practices so much a part of their culture that the discipline has been named after them (*disciplina Etrusca* or *auguralis*). Different phenomena and objects were used as media to ascertain the desires of the gods (regular and irregular celestial events, lightning, fire, and earthquakes, the shape or utterances of animals, flights of birds, movements of serpents, barking of dogs, forms of liver or entrails). Both in Etruria and Rome a numerous and well-organized hierarchy of functionaries existed for the practice of the sacred arts. (1958, p. 111 in 1961 edition)

Indeed, diffused through the Roman world, many of these techniques passed into medieval and modern culture.

**Diviner and doctor.** Callaway's account (1868–1870) of the combined divinatory and curative seance in Zululand emphasizes the close relationship believed to hold in many preliterate societies between the functions of divination and therapy. Sometimes, as in the case cited, the diviner and "doctor" are the same person, but more often the roles are specialized and performed by different individuals. Modern therapy is taking increasingly into account the psychosomatic character of many maladies and the importance of sociological factors in their etiology. In most preliterate societies bodily symptoms are regarded as signs that the soul or life principle of the patient is under attack or has been abstracted by spiritual forces or beings. Furthermore, it is widely held that these attacks are motivated by animosities provoked by breaches of cultural, mainly religious, prescriptions and/or by breaches of social norms regarded as binding on members of kin groups or local communities. Thus, to acquire a comprehensive understanding of why and how a patient was afflicted with certain symptoms by a spirit or witch, primitives seek out a

diviner who will disclose the secret antagonisms in social relations or the perhaps unconscious neglect of ritual rules (always a threat to the cultural order) that incited mystical retribution or malice. The diviner is a "diagnostician" who refers his clients to his colleague, the doctor or "therapist." The doctor in question has both shamanistic and priestly attributes. The division of labor which in more complex societies segregates and institutionalizes the functions of priest and medical man has hardly begun to make its influence felt. The diviner-doctor dichotomy does not depend, as does the priest-shaman dichotomy, upon contrasting roles in regard to the transhuman realm but upon different phases in a social process which involves total human phenomena—integral personalities, many psychosomatic complexes, multiple social relationships, and multiform communities.

**Modes of religious specialization.** As the scale and complexity of society increase and the division of labor develops, so too does the degree of religious specialization. This process accompanies a contraction in the domain of religion in social life. As Durkheim stated with typical creative exaggeration in his *Division of Labor in Society* ([1893] 1960, p. 169): "Originally [religion] pervades everything; everything social is religious; the two words are synonymous. Then, little by little, political, economic, scientific functions free themselves from the religious function, constitute themselves apart and take on a more and more acknowledged temporal character."

**Simple societies.** In the simplest societies every adult has some religious functions and the elders have most; as their capacity to hunt or garden wanes, their priestlike role comes into ever greater prominence. Women tend to receive more recognition and scope as religious functionaries than in more developed societies. There is some tendency toward religious specialization in such societies, based on a variety of attributes, such as knowledge of herbalistic lore, skill in leechcraft, the capacity to enter a state of trance or dissociation, and sometimes physical handicap that compels a man or woman to find an alternative means of support to subsistence activities. (I have met several diviners in central Africa with maimed hands or amputated limbs.) But such specialization can hardly be defined, in the majority of cases, as more than part-time or even spare-time specialization. Michael Gelfand's description of the Shona *nganga*, variously translated in the ethnographic literature as "medicine man," "doctor," or "witch doctor," exemplifies the sociocultural situation of similar practitioners in very many preliterate societies (1964).



The Shona *nganga* is at once a herbalist, a medium, and also a diviner who, possessed by a spirit of a dead relative, diagnoses both the cause of illness and of death. Yet, reports Gelfand,

when he is not engaged in his medical practice he leads exactly the same life as the other men of his village. He cultivates his land, looks after his cattle, repairs his huts, makes blankets or other equipment needed by his family. And the same applies to a woman *nganga*, who busies herself with the tasks expected of every Shona woman. . . . The amount the *nganga* does in his village depends, of course, on the demands of his patients, but on the average he has a fair amount of spare time . . . a fair guess would be [that there is a *nganga*] to every 800 to 1,000 persons. . . . The *nganga* is given no special status in his village, his chances of being appointed headman are the same as anyone else's. (1964, pp. 22-23)

*Complex societies.* To bring out best the effects of increase in scale and the division of labor it is necessary to examine religious systems at the opposite end of the gradient of complexity. Religion no longer pervades all social domains; it is limited to its own domain. Furthermore, it has acquired a contractual and associational character; people may choose both the form and extent of their religious participation or may opt out of any affiliation. On the other hand, within each religious group a considerable amount of specialization has taken place. Much of this has been on the organizational level. Processes of bureaucratization, involving rationality in decision making, relative impersonality in social relations, routinization of tasks, and a hierarchy of authority and function, have produced a large number of types, grades, and ranks of religious specialists in all the major religious systems.

For example, the Catholic clerical hierarchy may be considered as (1) the hierarchy of order, whose powers are exercised in worship and in the administration of the sacraments, and (2) as the hierarchy of jurisdiction, whose power is over the members of the church. Within the hierarchy of jurisdiction alone we find such manifold statuses as pope and bishop (which are held to be of divine institution); cardinal, patriarch, exarch, and primate (whose powers are derived by delegation expressed or implied from the holy see); metropolitan and archbishop (who derive their powers from their patriarch, exarch, or primate); archdeacon, vicar general, vicar forane, rural dean, pastor, and rector (who derive their powers from their diocesan bishop).

In addition to the clerical hierarchy there are in the Catholic church numerous institutes of the

religious, that is, societies of men and women approved by ecclesiastical superiors, in which the members in conformity with the special laws of their association take vows, perpetual or temporary, and by this means aspire to religious perfection. This is defined as "the heroic exercise of the virtue of supernatural charity" and is pursued by voluntary maintenance of the vows of poverty, chastity, and obedience, by ascetical practices, through charitable works, such as care of the poor, sick, aged, and mentally handicapped, and by contemplative techniques, such as prayer. Within each religious institution or congregation there is a marked division of function and gradation of office.

Thus there are many differences of religious status, rank, and function in a developed religious system such as the Catholic church. Differences in charismata are also recognized in such terms as "contemplative," "ascetic," "mystic," "preacher," "teacher," "administrator." These gifts may appear in any of the major divisions of the church: among clergy or laity, among hermits, monks, or friars, among female as well as male religious. Certain of these charismata are institutionalized and constitute the devotional pattern particular to certain religious institutions: thus there are "contemplative orders," "friars preachers," and the like.

*Medium-scale societies.* Other developed religions, churches, sects, cults, and religious movements exhibit degrees of bureaucratic organization and specialization of role and function. Between the situational specialization of religious activities found in small-scale societies and the full-time and manifold specialization in large-scale societies falls a wide variety of intermediate types. A characteristic religious dichotomy is found in many of the larger, politically centralized societies of west and east Africa, Asia, Polynesia, and pre-Columbian Central and South America. National and tribal gods are worshiped in the larger towns, and minor deities, daemons, and ancestral shades are venerated in the villages. At the village level we find once more the multifunctional religious practitioner. But where there are national gods there are usually national priests, their official servants, and worship tends to take place in temples or at fixed and elaborate shrines. Parrinder writes:

In the cults of the West African gods [for example, in Dahomey, Yoruba, and Ashanti] there are priests who are highly trained to do their work. These priests are often set aside from birth, or they may be called to the service of the god by being possessed by his spirit. They will then retire from their families and public life, and submit to the training of an older priest. The training normally lasts several years, during which

time the novice has to apply himself to learn all the secrets of consulting and serving the god. The training of a priest is an arduous matter. . . . [He] has to observe chastity and strict taboos of food and actions. He frequently has to sleep on a hard floor, have insufficient food, and learn to bear hardship. He is regarded as married to the god, though later he may take a wife. Like an Indian devotee, he seeks by self-discipline to train himself to hear the voice of his god. He learns the ritual and dances appropriate to the cult, receives instruction in the laws and taboos of the god, and gains some knowledge of magical medicines. (1954, pp. 100-101)

In these west African cults of deities there is a formal division of function between priests and mediums. In general, priests control mediums and carefully regulate their experience of possession. This situation is one solution to the perennial problem posed for priesthoods by what Ronald Knox (1950) has termed "enthusiasm," that is, the notion that one can become possessed by or identified with a god or God and that one's consequent acts and words are divinely inspired, even if they transgress religious or secular laws. In Dahomey, for example (Herskovits 1938), there are communal training centers, called cult houses or "convents," for mediums and assistants to priests. Here the novices are secluded for considerable periods of time. Part of their training involves the attempt to induce the return of the initial spirit possession that marked their calling. They learn later to produce coherent messages in a state of trance. During this period they are under the surveillance of priests. The Catholic church has similarly brought under its control as members of contemplative orders mystics and visionaries who claim "experiential knowledge of God's presence."

**Religious and political specialization.** In many primitive societies an intimate connection exists between religion and politics. If by politics we denote those behavioral processes of resolution of conflict between the common good and the interests of groups by the use of or struggle for power, then religion in such societies is pragmatically connected with the maintenance of those values and norms expressing the common good and preventing the undue exercise of power. In centralized political systems that have kings and chiefs, these dignitaries themselves have priestly functions; in many parts of Africa, for example, they take charge of observances which safeguard many of the basic needs of existence, such as rain making, sowing, and harvest rites, ritual to promote the fertility of men, domestic and wild animals, and so on. On the other hand, even where this is the case, there are frequently other specialized religious functionaries

whose duties are bound up with the office of kingship. An illustration of this occurs among the Bemba of Zambia, where the *Bakabilo*

are in charge of ceremonies at the sacred relic shrines and take possession of the *babanye* when the chief dies. They alone can purify the chief from the defilement of sex intercourse so that he is able to enter his relic shrine and perform the necessary rites there. They are in complete charge of the accession ceremonies of the paramount and the bigger territorial chiefs, and some of their number are described as *bafingo*, or hereditary buriers of the chief. Besides this, each individual *mukabilo* has his own small ritual duty or privilege, such as lighting the sacred fire, or forging the blade of the hoe that is to dig the foundations of the new capital. (Richards 1940, p. 109 in 1955 edition)

The *Bakabilo* constitute a council that exerts a check on the paramount's power, since the members are hereditary officials and cannot be removed at will. They are immune to the paramount's anger and can block the implementation of decisions that they consider to be detrimental to the interests of the Bemba people by refusing to perform the ritual functions that are necessary to the exercise of his office. A priesthood of this type thus forms a constituent part of the interior structure of the government of a primitive state.

In stateless societies in Africa and elsewhere, incumbents of certain ritual positions have similar functions in the maintenance of order and the resolution of conflict. The "leopard-skin chief" or "priest of the earth" (as this specialist has been variously called) among the Nuer of the Nilotic Sudan is a person whose ritual relationship with the earth gives him power to bless or curse, to cleanse a killer from the pollution of bloodshed, and, most important, to perform the rites of reconciliation between persons who are ready to terminate a blood feud. A similar role is performed by the "masters of the fishing spear" among the Dinka and the *tendaanas*, or earth priests, among the Tallensi and their congeners in the northern territories of Ghana. Similar religious functionaries are found in many other regions of Africa. They serve to reduce, if not to resolve, conflict within the society. As against sectional and factional interests they posit the commonweal. In these contexts, moreover, the commonweal is regarded as part of the cosmic order; breach, therefore, is mystically punished. The religious specialists are accorded the function of restoring the right relation that should obtain between society, the cosmos, and the deities or ancestral shades.

VICTOR W. TURNER



## BIBLIOGRAPHY

- BUBER, MARTIN (1936) 1958 *I and Thou*. 2d ed. New York: Scribner. → First published in German.
- CALLAWAY, HENRY (1868-1870) 1885 *The Religious System of the Amazulu*. Folk-lore Society Publication No. 15. London: Trübner.
- DURKHEIM, ÉMILE (1893) 1960 *The Division of Labor in Society*. Glencoe, Ill.: Free Press. → First published as *De la division du travail social*.
- ELWIN, VERRIER 1955 *The Religion of an Indian Tribe*. Bombay: Oxford Univ. Press.
- EVANS-PRITCHARD, E. E. (1949) 1954 *The Sanusi of Cyrenaica*. Oxford: Clarendon.
- EVANS-PRITCHARD, E. E. (1956) 1962 *Nuer Religion*. Oxford: Clarendon.
- FIRTH, R. W. 1964a Shaman. Pages 638-639 in Julius Gould and William L. Kolb (editors), *A Dictionary of the Social Sciences*. New York: Free Press.
- FIRTH, R. W. 1964b Spirit Mediumship. Page 689 in Julius Gould and William L. Kolb (editors), *A Dictionary of the Social Sciences*. New York: Free Press.
- GELFAND, MICHAEL 1964 *Witch Doctor: Traditional Medicine Man of Rhodesia*. London: Harvill.
- HERSKOVITS, MELVILLE J. 1938 *Dahomey: An Ancient West African Kingdom*. 2 vols. New York: Augustin.
- HOWELLS, WILLIAM W. 1948 *The Heathens: Primitive Man and His Religions*. Garden City, N.Y.: Doubleday. → A paperback edition was published in 1962.
- KNOX, RONALD A. 1950 *Enthusiasm: A Chapter in the History of Religion; With Special Reference to the XVII and XVIII Centuries*. New York: Oxford Univ. Press.
- LESSA, WILLIAM A.; and VOGT, EVON Z. (editors) (1958) 1965 *Reader in Comparative Religion: An Anthropological Approach*. 2d ed. New York: Harper. → See especially the unsigned introduction to Henry Callaway's article on p. 340.
- LOWIE, ROBERT H. 1954 *Indians of the Plains*. American Museum of Natural History, Anthropological Handbook No. 1. New York: McGraw-Hill.
- NADEL, SIEGFRIED F. 1954 *Nupe Religion*. London: Routledge.
- PARRINDER, EDWARD G. 1954 *African Traditional Religion*. London: Hutchinson's University Library.
- PARSONS, TALCOTT 1963 Introduction. In Max Weber, *The Sociology of Religion*. Boston: Beacon. → A paperback edition was published in 1964.
- PIDDINGTON, RALPH 1950 *Introduction to Social Anthropology*. 2 vols. New York: Praeger. → See especially Volume 1.
- RICHARDS, AUDREY I. (1940) 1961 *The Political System of the Bemba Tribe: Northeastern Rhodesia*. Pages 83-120 in Meyer Fortes and E. E. Evans-Pritchard (editors), *African Political Systems*. Oxford Univ. Press.
- WACH, JOACHIM 1958 *The Comparative Study of Religions*. New York: Columbia Univ. Press. → A paperback edition was published in 1961.
- WEBER, MAX (1922) 1963 *The Sociology of Religion*. Boston: Beacon. → First published in German. A paperback edition was published in 1964.
- WORSLEY, P. M. 1957a Millenarian Movements in Melanesia. *Rhodes-Livingstone Journal* 21:18-31.
- WORSLEY, P. M. 1957b *The Trumpet Shall Sound: A Study of "Cargo" Cults in Melanesia*. London: MacGibbon & Kee.

## II

## SOCIOLOGICAL STUDY

A religious specialist is a person in possession of ritual authority, esoteric knowledge, or spiritual gifts who is recognized as competent in the solution of religious needs. These needs may emerge from a demand for services on the part of a local group, from the exigencies of an organized cult, or from the political requirements of a ruling elite. The shaman witch doctor among the Ostyaks, the Brazilian Bororos, or the early American tribesmen is, from this perspective, as much a religious specialist as the Brahman priest or the Jewish rabbi.

Modern industrial societies, distinguished by cultural pluralism, status mobility, and rigorous performance standards, generate a wide range of religious needs and, thereby, divergent lines of religious specialization. Yet, in broad terms, religious specialization takes place within two spheres: within the professional clergy, who constitute the specialists for the organized religious bodies; and within the residual category of the grass-roots, client-oriented practitioners. Members of this latter category, including faith healers, clairvoyants, mediums, and self-styled prophets, acquire leadership and influence by their capacities to identify, articulate, and relieve spiritual problems. These practitioners, with local followings and *ad hoc* ritual techniques, occupy a distinct place in the realm of religious leadership today and wield considerable influence among the dispossessed who live outside the central orbit of middle-class life.

The professional clergy constitute an important subcategory of modern religious specialization. In turn, the term "clergy" bears two connotations—one technical and limited, the other descriptive and sociological. In the first sense, "clergy" (derived from the Greek term *kleros*, meaning "lot" or "portion") refers to the ordained ministers of the Christian religion. Inherent in this definition is the ritually conferred status of authority and responsibility in relation to a body of believers, or the laity. Within Roman Catholicism the clergy make up the major orders, with three canonically defined ranks: bishop, priest, and deacon. Priests, in turn, are either religious or diocesan (secular); the former live in a community under rule, whereas the latter work under the diocesan bishop as parish priests, attending to the care of souls. Protestant groups equate "clergy" with "minister," emphasizing the role of servitor, as opposed to the Roman Catholics, who place emphasis on sacramental function (Knox 1956, pp. 1-3). In the churches of the Anglican

Communion, however, the Roman Catholic terminology is generally applied (Hardy 1956, pp. 150–155). Hence, the connotation of a consecrated full-time professional leader is applicable throughout Christianity, even though differences are recognized according to tradition and theological position. Informal terms are also used to identify levels and types of Christian clergy: “higher” and “lower” clergy, terms used in the Eastern Orthodox church in prerevolutionary Russia; the “foreign” and the “indigenous” clergy, terms developing out of the missionary situation over the past century; “established” and “free” clergy, terms frequently used in countries like England and Sweden, where a state religion with a salaried clergy coincides with a nonconformist tradition. Characteristic divisions are made in terms of sexual status or family involvement—the celibate versus the married—and in terms of function—the curate, the priest, the preacher, the evangelist, the administrator, the therapist.

The second, more general use of the term “clergy” is increasingly employed by social scientists, journalists, and comparative historians to refer to all types of sacred experts. In this sense, the term identifies any full-time functionary connected with one of the major world religions. On this level, “the clergy” indicates a professional stratum or occupational group in society. “The American clergy” is a social category that automatically includes Protestant, Roman Catholic, Eastern Orthodox, and Jewish leaders.

Variations in the directions and types of specialization by function provide crosscutting bases for classification that cover both the professional clergy and folk experts. The roles of rabbi, the Islamic ulema, and the Roman Catholic canon jurist are specialties developed in connection with the interpretation of religious codes, holy laws, and ecclesiastical rules. Roles dealing with problems of illness within a religious framework include that of the exorcist, the pastoral counselor, and the Christian Science practitioner. Denominational officials, district superintendents, and seminary presidents hold key administrative-supervisory positions. Whether or not these same distinctions warrant inferences about elite status depends on the particular sociopolitical context.

### Classic studies of religion

Certain substantive issues, including priestly division of labor, sequential developments of religious types, and the functions of religious specialists in group life, have taken on special importance

in scholarly interests over the past hundred years. The discipline of comparative religion, launched in the mid-nineteenth century by philologists, historians, and students of mythology, produced a rich deposit of descriptive materials on the religions of ancient civilizations, including Babylonia, Egypt, Israel, China, Greece, and India. Although these scholars concentrated primarily on beliefs, attributes of deities, and ritual practices, they also contributed abundant information on priesthoods and other types of religious specialties, as well as the dynamics of priestly power and political change. However, this phase of early research, while still valuable, did not yield analytical arguments or explanatory theories regarding religious specialization.

The social evolutionists went to the other extreme. All types of religious and social phenomena were scrutinized and ordered into developmental phases, and speculative theory took precedence over comprehensive description. This enterprise made three contributions to the topic of religious leadership: the correlation of religious leadership forms with stages of cultural growth; the analysis of the role of psychological factors in the achievement of superior religious status; and the description of evolutionary development of specific religious types, with particular attention to the transition from the magician to the priest. Comte's work, although infused with the subjective optimism of the nineteenth-century positivists, connects evolutionary theory with the conditions for the rise of priesthood and its role in stabilizing the so-called theological stage of civilization.

The originators of the functional approach to religion and society—Robertson Smith, Émile Durkheim, Bronislaw Malinowski—did not identify religious leadership as a distinct theoretical problem, although a series of new concepts and generalizations on this category of specialization emerged indirectly from their work. Centering their primary attention on the universal properties of religion as a social phenomenon and its role in fastening group needs to the moral order, the functionalists considered ceremonial leaders and priestly activities secondary to religion as a key subsystem. In their dissatisfaction with earlier rationalistic and psychological views, these functionalists tried to demonstrate that religion is a distinct category of social experience; that it develops significant interdependencies with other institutional realms; and that it plays an undisputed function in society's operation and development. The combination of this broad theoretical polemic with their reliance



on materials from early and preliterate societies thereby led them to de-emphasize questions having to do with the internal differentiation of religious systems and the specifics of varying lines of ritual activity.

Freud's work holds a particular place in the intellectual development of this field. His theory of religion had three emphases: the origins of religion, centering around totemism; the psychological and cultural determinants of religious leadership; and the functions of religious action and beliefs for the personality system. Freud defined religion as a group neurosis—a fixation stemming from unresolved infantile conflicts that originate in early relationships with the parents, especially the father. The priest or religious leader serves as a surrogate parent, allowing for the neurotic expression of adult dependency needs but at the same time creating an unconscious ambivalence controlled through the observance of taboos. Furthermore, the presence of the religious leader re-creates the bonds of pseudo familism between member and member, as well as between member and leader. Although Freud provided a theoretical explanation for the emergence and persistence of religious leadership as a universal, he made no attempt to account for the development of special types of religious roles or to explain differences and variations from one society to another.

Max Weber provided the necessary breakthrough from historical-descriptive investigation and evolutionary theory. His comparative study of world religions, which focused on a theory of religious ideas and economic change, brought attention directly to religious functionaries as agents of social transformation or as conservators of tradition (Weber 1922). Priestly activity, historically combining with prophetic developments, produces a progressive "rationalization" of religious ideas. Weber saw these systematic orderings of supernatural conceptions as critical for levering men away from traditional, magical, and undifferentiated religious world views; hence he viewed the priesthood as a powerful educating force, which mediates between a world of magic and a world of ethical rationality. Utilizing the evolutionary technique, Weber delineated the development of various forms of religious specialization in sequences representing magician, mystagogue, and priest. Moreover, he compared the conditions under which different types emerge—prophet, lawgiver, teacher, or contemplative mystic—and went on to trace the economic, intellectual and political consequences of these various types of leadership and authority.

### Contemporary research

Since 1920 studies of religious specialists have moved away from the classic preoccupations with typologies and evolutionary sequences to one or another of the following problem areas: the clergy's personal and social characteristics, with emphasis on origins, recruitment factors, theological and political attitudes, and conceptions of their status (for example, Donovan 1958; Smith & Sjöberg 1961); the institutional context of clergy behavior and performance built around concepts derived from role theory and occupational issues, thus stressing status dilemmas, role strains, career lines, and the clergy's mode of integration with other professionals and clients in special settings, such as therapy networks and chaplaincies (Fichter 1954; Blizzard 1956; Cumming & Harrington 1963); broad diagnostic analyses of the clergy as a modern profession, with particular attention to the declining role of the sacred in industrial-urban life, the special criteria underlying the clergy's claim to professional status, and the sources of the clergy's alienation in the contemporary world (Hagstrom 1957; Fichter 1961; Gustafson 1963); and finally, comparative investigations of the clergy as members of local and national elite power structures, including their place among community influentials and as agents of change or tradition in the developing countries (Form & D'Antonio 1959; Underwood 1957; Pope 1942).

**Recruitment of religious specialists.** Clerical recruitment receives special attention from researchers because of the traditional religious leader's strategic role in communal life and his association with privilege, prestige, and even wealth. Criteria of eligibility for these elite positions are usually distinctive. Cross-culturally, the prerequisites range from physical fitness and extraordinary psychic capacities to hereditary status, sex, age, and spiritual experience. In the modern societies of the West, first attention is given to qualifications based on formal training, to the candidate's conviction of having received a "call," or to the seminary professors' judgments about the future clergyman's social skills, his psychological stability, his strength of professional commitment and, if he is Protestant or Jew and married, to the suitability of his wife for the pastoral situation. Contrasts in recruitment criteria are frequently overemphasized, to the neglect of examining the criteria's implications for other issues, such as the relationship between supply and demand.

Determination of eligibility by birth, which pro-

vides candidates without reference to need, may lead to oversupply. In the interest of making use of these resources, the division of labor involved in ritual is sometimes elaborated, allowing for the participation of greater numbers; such appears to be the case in Hinduism, where sacrifice requires the attendance of six or seven Brahman priests. Hereditary rights to sacred privilege thereby create vested interests in achieving and maintaining full ritual employment. However, different factors work to curtail this kind of elaboration. For example, believers can refuse to support additional religious leaders or turn to less costly substitute functionaries. The establishment of a centrally controlled ritual altar also reduces opportunities for leadership; thus members of the local hereditary priestly groups in ancient Israel lost their positions when sacrifice was limited to the temple in Jerusalem. In contemporary India, priestly Lingayat lineages solve the problem of oversupply by dividing the ritual role among the eligible by a process of rotation, each man holding the position of priest for one year in eight or ten.

Shortage, however, is a more serious problem than oversupply. As established sources decline or as emerging needs require additional recruits, religious bodies are forced to devise new recruitment strategies. Shortage problems of this kind are more characteristic of universalistic, missionary religions of the West, where the religious professions are of relatively low prestige, while the core spiritual functions remain the prerogative of an ordained elite. Roman Catholicism is often confronted with a particularly difficult dilemma in this respect: since salvation is dependent on the sacraments, and their administration is largely confined to the priest, large memberships in strategic regions are often without adequate pastoral care. This pattern is today particularly conspicuous in Latin America, where shortages of clergy reach crisis proportions. Not only is it common to have one priest for eight to ten thousand members, but the laity may be so distributed geographically as to make regular services impossible (Alonso et al. 1962; Pérez & Wust 1961). In order to mitigate the shortage, priests are being transferred from other nations, laity roles have been heightened, and the diaconate is being re-established as a part of the orders. Protestant bodies in Great Britain and the United States today also face problems of shortage. Particular concern is shown in the United States for recruiting clergy who know and understand urban issues. Most ministers are drawn from towns and small cities; yet the crux of contemporary Protestantism's problem

lies in developing influence in urban areas. Hence, the issue of a shortage of clergy is complicated by the issue of cultural background. Furthermore, while the need for clerical strength increases, the secular professions steadily offer wider opportunities and higher material rewards for college graduates. This intersection of rising demands and lowered supply places the churches in a double bind, with no immediate amelioration in sight.

**Training of religious specialists.** The variations found in procedures for training the clergy and other religious specialists are striking and may be viewed in terms of structural context, content, and duration. The principal differences in structural context are made plain by contrasting the father-son pattern that is found among traditional priestly family groups with the autonomous, interdenominational, university-affiliated seminaries that provide training for future ministers in the major American Protestant denominations. Between these two poles fall the teacher-disciple relationship, the work of seminaries sponsored and controlled by particular religious bodies, and the special schooling required for candidates to the religious orders and religious congregations. In the early Roman Catholic church, priests were trained in cathedral schools, which generally gave way to the universities in the twelfth and thirteenth centuries. The Council of Trent in the sixteenth century established the requirement of seminary training. Prior to the 1917 revolution, Russian clergy of the Eastern Orthodox church were trained in seminaries, but there were only one or two of these, and most of the lower clergy were not trained at all. The center of scholarly training for the Islamic ulema is al-Azhar in Cairo; however, throughout most of the world, training in the Koran is carried out by local teachers, who pass the tradition on to students. In the United States the need for additional Protestant preachers and ministers in the eighteenth and nineteenth centuries led to the founding of numerous religious colleges. Moreover, an increase in the demand for licensed clergy in modern times has created a flock of diploma mills and pseudo seminaries.

Training does not always prepare religious specialists for their work. Because this holds true for the major Christian bodies today, a critical evaluation of the seminary and its curriculum is taking place. The clergy are disciplined in church history, theology, and dogma only to find that the pastoral role demands administrative capabilities, human relations skills, practical techniques in the programming of secular activities, and the ability to eval-



uate and act in relation to community needs. This confusion between professional expectations and professional realities is held to be the key source of low morale and insecurity among the clergy (Niebuhr 1956).

**Career mobility.** Religious specialists are career men in that their commitment to the sacred office is usually for life. Thus, patterns of placement and tenure are understandably important. Many Anglican clergymen, for example, hold beneficed positions from the time of their ordination until their death. This tenure system has many drawbacks, not the least being the episcopate's loss of control over the lower clergy. The Roman Catholic secular priest is usually assigned in the diocese within which he has received seminary training. In the congregationally centered Protestant denominations, the ordained minister is contracted for by the local church, though frequently on the basis of recommendations from denominational officials. This same policy holds in Conservative Judaism. Procedures of appointment, whether from above or below, affect the clergyman's mobility and his conceptions of success. In the congregational churches, ministers move frequently and widely. Competition is keen for the best churches.

Barriers to advancement are of many types, including particularistic preferences for manner or speech as well as slow turnover at the top levels. The growth of administrative positions and the opening of special ministries in the big cities are helping to relieve some of these mobility pressures. When openings do exist, the criteria for promotion are often unclear. The close involvement of the clergy in ritual duties, symbolic acts, and ministries of love complicates the whole process of evaluation. Lacking explicit criteria for judging clerical performance, many religions fall back on the seniority principle. The clustering of the older men at the top tends, of course, to increase the inflexibility and conservatism that are often associated with religious organizations. Although the clergyman's position in society is a secure one, the internal pattern of his professional success often depends on skillful maneuvering and patience rather than on accomplished performance. In particular cases success is not anticipated, at least not in terms of moving up. The lower clergy in the Russian church before the Revolution were prohibited from receiving appointments to vacant sees, both because they were married and because they were socially inferior. This barrier between the church's higher offices and the local clergy has been maintained throughout most of the past four hundred years in Latin America.

**Political position of the religious elite.** The potential conflict between priests and rulers, or between religious and secular elites, erupts from various causes, revealing itself in at least three situations: when sacred principles are used to challenge the legitimacy of political authority, as most clearly exemplified in the contest between *regnum* and *sacerdotium* in western Europe from the eleventh to the fourteenth century (Parker 1955); when religious leaders gain a monopoly over a sector of public life that is deemed essential to the development of society, as was the case in the conflict that emerged in France during the 1880s over the secularization of the schools (Moody et al. 1953); and when religious leaders or sacred authorities initiate programs that violate moral sensibilities or secular law, for example, the Mormons' practice of polygamy in the United States after their move to Utah. Conflict is least likely when the religious cult is controlled by the government, with the priests serving as official bureaucrats, and when, as in classical China, the state officials perform the ritual ceremonies. The complete separation of church and state is an equally stable solution, demonstrated by the history of the United States over the past 180 years.

The social strata from which the clergy are drawn is another relevant factor. If they are identified with a single class, as in prerevolutionary France, hostilities toward the clergy may polarize, drawing power from general social antagonisms. When, on the other hand, clergy are drawn from all classes, class hatred is potentially reduced, since every family or status group has a clerical representative. Religious elites gain adaptive advantages from other sources, including their capacity to draw personnel and finances from other countries—as in present-day Roman Catholicism—their possession of publishing houses and parochial schools, and the extent to which the religious system is centralized, allowing control over the movement and allocation of religious specialists.

Over a period of time, the position a religious elite holds in the political sphere may change because of changes in the elite's function and its relationship to levels of social structure. In the United States the power of the clergy on the community level is changing from that of decision maker to that of an agency for mobilizing support on issues already decided. Among the clergy in France, emphasis has moved away from competing for formal power on the national level toward the promoting of value change in local settings. In all modern industrial societies, the trend of the clergy is away from open political struggle and toward a

reliance on their religious authority to influence issues that are held to be of moral and social importance. In broad terms, the change is from open, direct maneuvers to indirect persuasion and suggestion, frequently backed up by joint statements that crosscut denominational and faith lines.

**Internal conflict.** Although overtly symbolizing solidarity and stability, religious elites are frequently divided by internal strife and competition. In the present era new factors are emerging to foment such differences. Developments in Protestant groups at the level of central administration, with their attendant changes in authority and power, infringe on the autonomy of the local pastors. Ecumenical moves and cross-denominational cooperation are stimulating reactions among clergy who take historical differences seriously. In another dimension, the problem of defining the mission of the church in society provides room for varied concepts and sharply defined differences of opinion: some groups see the church as an agency of social reform, whereas others view the congregational unit as a center of religious fellowship, ritual worship, and inwardly focused spirituality. Roman Catholic divisions today center less on matters of belief and dogma than on conceptions of the church in relation to society and on the nature of authority and its appropriate means of institutionalization. The major issues in these debates center on the opposition between the ascetic view of the church and the view emphasizing the church incarnate; between the conservative view of priestly authority and the view favoring lay-clergy cooperation; between bishops favoring the supremacy of the pope and those holding for increased decentralization; and finally, between those regarding religious orders as fortresses of spiritual contemplation and those desiring their increased role in the active apostolate.

**Religious leaders and the laity.** Many of the vital issues underlying the problems and dilemmas of religious leaders emerge in the relationship of the clergy to laymen. In the past, conflicts between laymen and clergy were embodied in the laity's rebellion against sacred privilege, lax morality among priests, the economic burden of church taxes, and the methods of clerical appointment. Different problems have emerged in the modern period to foster clergy-laity tensions. The enormous rise in educational level of church members in Europe and the United States robs the religious leader of his earlier position as a man of learning; at the same time, the laity demand more from him as minister, social leader, and administrator. Matters of control over ecclesiastical property and ap-

pointment prerogatives are now less central, in most instances, than that of the social relationship between the pastor and his charges. Although the norms governing this basic relationship are in a process of change, they tend to give priority to expectations of warmth, personalism, and service. The full-time religious leader is expected to live up to these norms and to be accessible and available at all times. The core problem that emerges from this typical situation has its source in the religious leader's struggle to maintain effective distance from the demands and invitations that laymen consider appropriate. Various arrangements help to offset these difficulties: closed residence for priests; location of the minister's home at a distance from the house of worship; segregated religious assemblies, such as special retreats for religious leaders; distinctions in dress; and limitation of contact to ritual occasions or to special appointment hours. Centralized religions often follow a policy of periodically reassigning their local leaders, thereby indirectly reducing patterns of long-term involvement between leader and membership.

A distinctive and growing source of clergy-laity conflict is in evidence in the contemporary Roman Catholic church. The movement toward increased lay participation, initiated by Leo XIII in the last decades of the nineteenth century, has gradually taken root in France, Belgium, the United States, Brazil, and Chile. For the past seventy years the church has tried to reach alienated membership groups, ranging from the French urban proletariat to the Chilean rural peasantry. This ambitious undertaking places an enormous strain on church labor reserves, and local clergy are inadequate to carry out this new responsibility. The laity has therefore become a potential auxiliary work force, assuming organizational assignments such as teaching the catechism, actively proselytizing, organizing youth groups, and assisting with the administration of the sacraments (Marie Joseph 1953). These changes provoke tensions between the diocesan clergy and the laity, since the clergy's power and prestige tend to decline, while the status of the layman rises. Thus, the membership both expect and demand greater voice in local church administration and policy decisions. It is clear that the priest's "charisma" of office is no longer an adequate basis for his authority. He must be knowledgeable as well as schooled, and the power of his leadership increasingly rests on his spiritual-social qualities rather than on the symbolic duties of the sacramental role.

**Religious elites and the social order.** A view of the broader role the clergy play in society under-



lines other aspects of religious leadership's complex nature. Except for some monastic communities and certain types of professional mystic, sacred men are always intimately bound up with the wider social order. The over-all trend of the clergy's place in society is that of narrowed functional importance, while the trend within the religious unit is toward increased scope of functional responsibility. A recognition of this dual change helps explain many of the current findings relative to the clergyman's confusion and insecurity. The interpretation that some sociologists give to the "minister's dilemma" holds that added administrative and social responsibilities in the local church have created strains in carrying out the roles of pastor, preacher, and ritual agent. Hence, multiple roles are taken as the cause of the clergyman's distress (Blizzard 1956; Fichter 1954).

This explanation is unsatisfactory, since it fails to take account of the changing place of the religious leader in society. The clergyman has often filled multiple roles. In seventeenth-century England, for example, the local clergyman served as lawyer, teacher, counselor, doctor, and community social leader, as well as preacher and administrator of the sacraments; his time and energies were more in demand than those of his modern counterpart. However, the important fact is that these multiple roles brought status and respect from society: he was visible, needed, and rewarded. Today the clergyman still finds himself with multiple roles, but they are bound up with the internal activities of the religious unit. The organized religions are now confronted with the painful problem of re-establishing a position of usefulness and influence in society. Organizational activities thereby assume strategic importance, as a means of attracting people and as a way of increasing leadership effectiveness. In this situation the local clergyman is pushed to become planner, strategist, organizer, and action leader. These demands, however, arise from within the religious system, not from society. He is thereby separated in his work from the eye of the public, the major spheres of social reward, and the growing prestige of the secular professions; yet he is expected to meet the requirements of a demanding position.

### The clergy and social change

Controversy continues to surround the clergy's relationship to social change. Priestly professions and conservatism often go hand in hand. At least three factors play a decisive role in building this conservatism: high priests and religious officials are often recruited from society's upper strata; re-

ligious beliefs characteristically stress otherworldly rewards, thus devaluating secular progress and accomplishment; and ritual exactness and adherence to form are assumed to hold a central role in maintaining the benefits and favors of supernatural entities. Thus, tampering with time-honored techniques may disturb an established positive balance. Certain conceptual refinements, as well as changing social conditions, serve to qualify this static picture. For example, typologies applicable to religious roles—such as "priest" versus "prophet," "pastor" versus "administrator," or "local" versus "cosmopolitan"—provide useful distinctions between specialists who focus on religion's conserving functions and those who attempt to create new breakthroughs in religious meaning, action, and organization (Wach 1944). Modern theologians symbolize this innovative side.

Moreover, within a given type of religious specialty or within a historical period, the institutional context of the specialist's work must be taken into account. Clergymen who are assigned duties outside the denominational-parish context are often more committed to innovation and change than their secular counterparts. With the proliferation of special ministries, including factory programs, university and hospital chaplaincies, and inner-city missions, a growing number of clergy are partly freed from traditional authority and conformity-based parish pressures. However, these new men are vulnerable to other pressures, which emerge in the outside settings, as the "worker priest" experiment showed (Edwards 1961; Poulat 1965). Ministers who lead local congregations, on the other hand, are tied directly to a traditionally defined religious unit, from which they draw their rewards. Their stand on a moral or sociopolitical issue—whether civil rights, ecumenicism, or federal aid to education—depends on the values of the dominant lay groups. Conservatism or its opposite therefore tends to be built into the situation. Present-day critics of the Southern clergy's reluctance to lead out for racial integration overlook this intimate relationship between context and behavior (Campbell & Pettigrew 1959).

Recent empirical studies suggest that the clergyman's effectiveness as a counselor, or as an agent of personality change, is also limited by this tie-in with a concrete religious group (Cumming & Harrington 1963). Although he is accessible, shares key values with his parishioners, and is often well prepared to understand the depth factors that produce personal stress, he lacks the professional distance of the psychiatrist or caseworker. Two consequences flow from this situation. First, though

the clergyman is frequently consulted before any other professional by members of his congregation with personal problems, the relationship with the "patient" is usually weak or easily interrupted. Members tend to withhold information, fearing that it may get "loose in the system," or they withdraw from the relationship altogether. Second, the clergy's identification with a religious tradition leads other professionals who give counsel or therapy (psychiatrist, social worker, doctor) to overlook the local clergyman as a member of the broader deviance-control system. Even when a patient's problem appears to warrant "religious counseling," referral to a minister is seldom made; yet the clergy steadily refer members of their congregation to secular professionals. Thus, the clergyman in his role as therapist gives more than he gets, not simply because he is less qualified but also because he is viewed by his secular colleagues in the community as an agent of a particular belief system, and more broadly, as a man with a "moral outlook."

The overseas missionary, also, is a distinct type of religious specialist. This fact calls for some extension of these qualifications about the clergy's role in social change. The clergy's fostering of schools, hospitals, agricultural technology, and programs of community development in Asia, Africa, and Latin America represents a major type of directed change (Maddox 1956). In fact, the missionary led the way in various types of socioeconomic reform throughout these areas, between the late nineteenth century and the end of World War II. Since 1945 government technical missions, United Nations teams, and special overseas programs, such as the Peace Corps, have largely eclipsed the missionary's frontier role in promoting change. Moreover, the implicit association often made between foreign missions and colonial interests has further reduced missionary opportunities. Paralleling these developments, leaders of the organized religions increasingly recognize that church survival and growth depend not only on overseas work but also on new types of entrepreneurship on the domestic front, especially in the metropolitan areas and among workers and racial and ethnic minorities. In those remaining areas where proselyting is permitted, foreign missionaries also confront growing pressures to turn their work over to an indigenous clergy. The Catholics appear to be more successful in this grass-roots endeavor than the Protestants, partly, of course, because of differences in administrative and financial policies.

The role of religious leadership in social change is further clarified by taking account of the re-

ligious system's authority structure. Centralized, hierarchically organized priesthoods frequently breed divisions of a "higher" and "lower" order, the latter stratum being held in subservience by the superiors. Tensions that emerge from this caste-like pattern are available for displacement into political directions during periods of social unrest, the lower clergy uniting with other oppressed groups as supporters of the liberal cause. This happened in Latin America from 1810 to 1830, and in Russia following the 1917 revolution. Hence, both tensions within religious leadership groups and conflict among religious elites may help foment social change and liberal activity, depending on how these antagonisms become fused with broader political and economic forces. In contemporary Latin America, especially in Chile and Brazil, the Roman Catholic clergy are demonstrating an unusually progressive position on social change (Vallier 1965). This is a surprise, considering the historical position of the church in these areas. Further examination reveals, however, that in both Chile and Brazil secular and religious threats are very real, one source being the radical political doctrines of the socialist and communist left, and the other the successes of the Protestant sects. Liberal steps, such as the distribution of church lands to peasants, the establishment of technical-training schools, and trade-union activity, are seen as new ways of capturing religious loyalties. Hence, to understand the position that priests or ministers take in fomenting or restricting social change, a whole range of referents need to be specified.

**New patterns of religious specialization.** Viewed from within the broader context of the modern professions, religious specialization in Christianity exhibits several noteworthy patterns. As in medicine, law, and university teaching, growing lines of differentiation (therapy, social research, ecumenicism, liturgy, and domestic missions) create pressures for integrative-coordinative roles. Thus, experts in administration, communication, and planning gain visibility and influence. In American Protestant denominations these coordinative problems loom very large, and internal power struggles increasingly emerge around these units. American Roman Catholicism, on the other hand, is well-developed as an administrative and policy-making body but lacks effective lines of specialization with respect to problems of the laity, interfaith cooperation, and the linking of group worship to communal forms other than those resulting from traditional ethnic bonds. In Europe the dominant line of specialization is the theologian's role within the realm of Judaeo-Christian ideas as a culture; theo-



logical activity is heavily linked to Biblical studies—a combination that is producing a major creative thrust. The core problems for European Christianity emerge at the level of meaning and ideology rather than in the spheres of organization and action. Thus, possibilities for the development of Christianity hinge on the theologians' capacities to establish competitive systems of faith vis-à-vis Marxism, existentialism, and secular totalitarian mystiques. In broad terms, lines of religious specialization differ not only by Protestant and Catholic traditions but also by continents.

### Directions of research

From the foregoing considerations, it is clear that the subject of modern religious specialists comprises diverse substantive problems and important analytical issues. Religious specialists not only continue to hold a distinctive place in the contemporary world but also make up a key category in the study of occupations, social movements, culture, and organization. Yet this field presently lacks vitality and direction within the social sciences. New evaluations are needed to restore to it some of the contributory capacity it earlier possessed. Several suggestions are in order.

One possibility is to connect the study of religious specialization more fully to the problem of the churches' ambitions to retain and extend forms of religious influence in modern society. Both Christian and non-Christian groups seek to maximize their spheres of dominance, and in these endeavors traditional strategies and former linkages with the social order lose importance. Consequently, innovation, entrepreneurship, and institution building, as instruments of religious influence, are highly significant and thus become points of structural growth and change. Comparisons are possible between major religious traditions—Christianity, Judaism, Islam, Buddhism, and Hinduism—and within distinct faiths (Vallier 1962). The three great modern adaptations of Roman Catholicism—in western Europe, the United States, and Latin America—constitute only one subspecies within which religious specialization and the problem of achieving new lines of religious control could be studied. The point, of course, is that religious specialists should be viewed as more than members of an isolated occupational category or local leaders frustrated by new pressures. An approach that focuses on patterns of specialization in relation to the evolution and change of total religious systems would be more fruitful.

A second possibility involves the comparative study of religious specialists as factions within a religion's "political" subsystem. Although conflicts

and power struggles are endemic to all religions, these divisions are too often examined within the limited "liberal" and "conservative" categories. This is a risky oversimplification. The liberal Roman Catholic development in modern Europe or in Latin America is actually made up of diverse interests, orientations, and ideologies, each with its own elite and identity. In this same sense the Roman Catholic hierarchy of the United States is often labeled "conservative" relative to the liberal groups in Europe. However, no one would seriously argue that the conservative elite in the American hierarchy hold the same views and positions as the Italian prelates who control the Curia. Liberal Protestant clergy are often lumped together by such flimsy criteria as their attitudes toward the Bible or their reading of the *Christian Century*, but again the spectrum is very wide. Moreover, to identify a segment that holds similar attitudes does not allow any inference about the role these individuals play in the internal political arena. In addition to the need for work on intrareligious conflicts and factions, further extensions are appropriate for systematically relating religious elites to nonreligious elites in the wider context of modern society. Studies of community power structures provide suggestive leads for this problem.

Other possibilities are open for developing the subject of religious leadership. The traditional emphasis on typological diversity tends to obscure the central question as to what sacred positions have in common. Distinctions by rank, function, and training are often taken to be more important than the characteristics that religious specialists share as a distinct category of social structure. The simple fact that all religious specialists are focused primarily on the realm of the nonempirical and religioethical absolutes sets them apart from statesmen, business executives, and doctors. The clergyman lacks a technical body of knowledge that can be applied to the solution of empirical problems. In that sense he cannot practice his profession; instead he must live it. For the most part, his base of knowledge is normative and closed rather than scientific and open. He interprets, while his colleagues in the secular professions prescribe; he tries to give meaning to the event, since he cannot control it. Thus, to place the clergy along with lawyers, academicians, social workers, and engineers into one broad category marked "professionals" is to violate several distinctives of singular theoretical importance. As the special properties of sacred elites become clarified, knowledge of religious dynamics and of the nature of the professions will be advanced.

[See also PSYCHIATRY, article on THE RELIGIO-PSYCH-  
ATRIC MOVEMENT; RELIGIOUS ORGANIZATION; SECTS  
AND CULTS. Other relevant material may be found in  
the biographies of COMTE; FREUD; WEBER, MAX.]

## BIBLIOGRAPHY

- Statistics on the modern clergy can be found in *Bilan du monde* 1964; U.S. Bureau of the Census 1964; the yearbook of the National Council of the Churches of Christ in the United States of America; The Catholic Directory 1960; *Estudios socio-religiosos latinoamericanos*. For bibliographies on religious specialists, including historical trends and special groups, consult Wach 1944; Niebuhr & Williams 1956; Salisbury 1964. On the clergy's training and its professional codes, see Fichter 1961. For the problems of the modern clergy, consult Vatican Council 1966; Poblete Barth 1965; Edwards 1961. Empirical studies of the modern clergy's political and social views are Lenski 1961 and Campbell & Pettigrew 1959.
- ALONSO, ISIDORO; POBLETE, RENATO; and GARRIDO, GINES 1962 *La iglesia en Chile: Estructuras eclesíásticas*. Fribourg (Switzerland): Oficina Internacional de Investigaciones Sociales de FERES.
- Bilan du monde*. 2 vols. Encyclopédie catholique du monde chrétien. 1964 Paris: Casterman.
- BLIZZARD, SAMUEL W. 1956 The Minister's Dilemma. *Christian Century* 73:508-510.
- CAMPBELL, ERNEST Q.; and PETTIGREW, THOMAS F. 1959 *Christians in Racial Crisis: A Study of Little Rock's Ministry*. Washington: Public Affairs Press.
- The Catholic Directory*. 1960 London: Burns & Oates.
- CONGAR, MARIE JOSEPH (1953) 1957 *Lay People in the Church*. Westminster, Md.: Newman. → Author's secular name is Georges Yves Congar. First published as *Jalons pour une théologie du laïc*.
- CUMMING, ELAINE; and HARRINGTON, CHARLES 1963 Clergyman as Counselor. *American Journal of Sociology* 69:234-243.
- DONOVAN, JOHN D. 1958 The American Catholic Hierarchy: A Social Profile. *American Catholic Sociological Review* 19:98-112.
- EDWARDS, DAVID L. (editor) 1961 *Priests and Workers: An Anglo-French Discussion*. London: Student Christian Movement Press.
- Estudios socio-religiosos latinoamericanos*. → Published since 1961.
- FICHTER, JOSEPH H. 1954 *Social Relations in the Urban Parish*. Univ. of Chicago Press.
- FICHTER, JOSEPH H. 1961 *Religion as an Occupation*. Univ. of Notre Dame (Ind.) Press.
- FORM, WILLIAM H.; and D'ANTONIO, WILLIAM V. 1959 Integration and Cleavage Among Community Influentials in Two Border Cities. *American Sociological Review* 24:804-814.
- GUSTAFSON, JAMES M. 1963 The Clergy in the United States. *Daedalus* 92:724-744.
- HAGSTROM, WARREN O. 1957 The Protestant Clergy as a Profession: Status and Prospects. *Berkeley Publications in Society and Institutions* 3:1-12. → Now called *Berkeley Journal of Sociology*.
- HARDY, EDWARD ROCHE JR. 1956 Priestly Ministries in the Modern Church. Pages 149-179 in H. Richard Niebuhr and Daniel D. Williams (editors), *The Ministry in Historical Perspectives*. New York: Harper.
- KNOX, JOHN 1956 The Ministry in the Primitive Church. Pages 1-26 in H. Richard Niebuhr and Daniel D. Williams (editors), *The Ministry in Historical Perspectives*. New York: Harper.
- LENSKI, GERHARD (1961) 1963 *The Religious Factor*. Rev. ed. Garden City, N.Y.: Doubleday. → See especially Chapter 7.
- MADDOX, JAMES G. 1956 *Technical Assistance by Religious Agencies in Latin America*. Univ. of Chicago Press.
- MOODY, JOSEPH N.; MICAUD, CHARLES A.; and VIGNAUX, PAUL 1953 *Catholicism and Society in France*. Pages 93-277 in Joseph N. Moody (editor), *Church and Society: Catholic Social and Political Thought and Movements, 1789-1950*. New York: Arts.
- NATIONAL COUNCIL OF THE CHURCHES OF CHRIST IN THE UNITED STATES OF AMERICA *Yearbook of American Churches*. → Published since 1916, under various titles. The National Council is the successor to the Federal Council of the Churches of Christ in America.
- NIEBUHR, H. RICHARD 1956 *The Purpose of the Church and Its Ministry: Reflections on the Aims of Theological Education*. New York: Harper.
- NIEBUHR, H. RICHARD; and WILLIAMS, DANIEL D. (editors) 1956 *The Ministry in Historical Perspectives*. New York: Harper.
- PARKER, T. M. 1955 *Christianity and the State in the Light of History*. New York: Harper.
- PÉREZ RAMÍREZ, GUSTAVO; and WUST, ISAAC 1961 *La iglesia en Colombia: Estructuras eclesíásticas*. Bogotá: Oficina Internacional de Investigaciones Sociales de FERES.
- POBLETE, RENATO 1965 *Crisis sacerdotal*. Santiago: Editorial del Pacífico.
- POPE, LISTON 1942 *Millhands and Preachers: A Study of Gastonia*. New Haven: Yale Univ. Press; Oxford Univ. Press.
- POULAT, ÉMILE 1965 *La naissance des prêtres ouvriers*. Paris: Casterman.
- SALISBURY, W. SEWARD 1964 *Religion in American Culture*. Homewood, Ill.: Dorsey. → See the bibliography on pages 503-504.
- SMITH, JAMES O.; and SJOBERG, GIDEON 1961 Origins and Career Patterns of Leading Protestant Clergymen. *Social Forces* 39:290-296.
- UNDERWOOD, KENNETH W. 1957 *Protestant and Catholic: Religious and Social Interaction in an Industrial Community*. Boston: Beacon.
- U.S. BUREAU OF THE CENSUS 1964 *Statistical Abstract of the United States*. Washington: Government Printing Office.
- VALLIER, IVAN 1962 Church, Society, and Labor Resources: An Intra-denominational Comparison. *American Journal of Sociology* 68:21-33.
- VALLIER, IVAN 1965 *Religious Elites in Latin America: Catholicism, Leadership and Social Change*. Unpublished manuscript.
- VATICAN COUNCIL, SECOND, 1962-1965 1966 *The Documents of Vatican II*. New York: America Press. → A paperback edition was published in the same year. See especially pages 396-429 on "Decree on the Bishops' Pastoral Office in the Church"; pages 437-457 on "Decree on Priestly Formation"; pages 466-482 on "Decree on the Appropriate Renewal of the Religious Life"; and pages 532-576 on "Decree on the Ministry and Life of Priests."
- WACH, JOACHIM (1944) 1958 *Sociology of Religion*. Univ. of Chicago Press. → See especially pages 331-374 on "Types of Religious Authority."
- WEBER, MAX (1922) 1963 *The Sociology of Religion*. Boston: Beacon. → First published in German.



## RENT

There are at least three distinct classes of meanings for *rent*: the common usage, the usage in classical economics, and the usage in modern economics. These are discussed in turn in this section. More restrictive meanings or special kinds of rent will be discussed under the appropriate general classification. In succeeding sections the determination of *rent* in its various senses will be discussed.

**In common usage.** As used in everyday speech, *rent* means a payment for the services of a material asset for a specific period of time. As such, rent is comparable to the wages of labor, and, as suggested by Frank Knight, both wages and rent might be classed under the heading *hire value*. The term *rental value* might also be used for what is commonly called rent to avoid confusion with the technical economic meaning of rent. Real estate, or land and buildings, is the type of asset most frequently rented, but a wide variety of other consumer and producer durable goods are rented as well.

Renting or leasing as an alternative to ownership may take place for a variety of reasons, one of which is infrequency or short duration of use relative to costs of purchase and sale. Houses or apartments are typically rented by smaller households, probably because small dwellings are relatively more costly than larger ones when provided in the form of single-family dwellings, and by younger families, who have little accumulated nonhuman wealth (i.e., wealth other than their ability to earn income with their own labor) and for whom the interest cost of capital invested in ownership is high. Business firms may choose leasing in preference to ownership for quite similar reasons. In addition, renting of any asset may be preferred by households or business firms that have a greater than average aversion to the risk of the change in value of assets that ownership entails.

The actual rental payment made by the user of an asset includes several components. First, gross rent, or rental value, includes a payment to cover depreciation, which is the decline in the value of the asset during its period of use. Second, where the owner rather than the user of the asset is liable for payment of property taxes, as is typically the case in the United States, the rental payment will include an allowance to cover these taxes. The rental payment, like an interest payment, may also include a component to cover the cost of making and servicing the lease agreement. And finally, the rent of any asset includes a payment to cover interest

costs on the value of the asset during its period of use.

There are a variety of special types of rental payments. The term "ground rent" refers to a payment for the use of land on a long-term lease, frequently 99 years. Such payments provide a relatively sure and very stable income stream, which some investors find attractive. "Space rent" in the consumer expenditure tables of the U.S. national income accounts refers to the rental value of a structure and the land on which it is situated, apart from any payment for utilities and furnishings supplied by the landlord. In the U.S. Census Bureau reports on housing, "contract rent" means the payment agreed on by tenant and landlord and includes any payment for utilities and furnishings supplied, while "gross rent" refers to contract rent plus the estimated value of all utilities not furnished the tenant by the landlord. Finally, the "rental income of persons" as a component of the U.S. national income accounts refers mainly to the income actually received directly by households from rented nonfarm property and the imputed net return from owner-occupied nonfarm housing. It also includes the net returns received by persons from farm real estate, but this component is small relative to the other two. It should be noted that the rental income of persons shown in the U.S. national income accounts includes only a small fraction of the rental value of land or of material assets generally. Most rental income is actually paid out in the form of corporate income or income to unincorporated businesses.

**In classical economics.** Economists realized quite early that much of what is commonly called rent is really nothing but the return to the various forms of capital investment. Thus, the classical economists reserved the term "rent" for a type of payment they believed to be qualitatively distinct from wages and interest. This was that part of the landlord's return for the use of "the original and indestructible powers of the soil," as Ricardo put it. Rent of land in this sense might arise because of the limitation on its quantity, as stressed by Malthus, and this type of rent was sometimes called "scarcity rent." Land rent might also result from differences in fertility and location. Rent arising from this source was frequently called "differential rent." It is easy to make too much of the distinction between scarcity and differential rent, however, since the most fertile or favorably located land would earn a differential return over other land only if the former were scarce. [See the *biographies* of MALTHUS; RICARDO.]

Ricardo especially stressed that the most fertile

and favorably situated land would be cultivated first. With the increase in population and wealth, according to Ricardo, land of inferior quality is taken into cultivation. The exchange value of raw produce would rise because more labor (and capital used with labor in fixed proportions, presumably) would be required to produce on the less fertile land. The first land would then yield a return equal to the difference between the produce obtained by the employment of equal quantities of capital and labor on each quality of land. Rent, Ricardo argued, does not enter into the cost of production. The latter is determined by the quantities of other factors employed at the margin of cultivation on land that yields no rent. "Corn [i.e., food] is not high because rent is paid, but a rent is paid because corn is high," in Ricardo's words.

Even in the more limited sense in which the term was used by the classical economists, it is difficult to ascribe the returns to land to some "original and indestructible" properties of nature. In a very real sense, land as a productive factor was created as capital by pioneers who gave up other income in traveling to and preparing the land for cultivation. To these pioneers, the return to their land after an allowance for the obvious costs of buildings and other visible improvements was merely a return to their earlier sacrifices. To later owners of the land, its return was merely a return to their fluid capital, which might alternatively have been used to acquire other material assets. Land today is reclaimed through irrigation and drainage, and its fertility is obviously capable of exhaustion. In addition, the supply of favorably located land is continually being increased by improvements in transportation. Finally, when land has more than one use, the maximum income it can earn in alternative uses is a part of the cost of production for the commodity it helps to produce. Thus, oranges may be expensive because of high returns to land used to grow oranges if the high returns to land upon which oranges are grown result from its high value for alternative residential and other urban uses. [See COST.]

Not only is land reproducible and destructible in much the same way as other material assets but other assets may possess properties similar to those the classical economists ascribed to land. Any productive factor, including labor, is limited in absolute amount and thus earns a return that results from its scarcity. Quality differentials exist for most classes of productive factors no matter how narrowly defined; differential incomes are earned by members of a particular class of productive factor as a result of their differing productivities.

And many productive factors earn incomes that exceed the minimum amount necessary to insure their employment in the specific use to which they are put.

In modern economics. For reasons such as these, economists today rarely think of rent as a qualitatively distinct type of payment accruing to a specific type of productive factor. Rather, "rent," along with the terms "cost" and "profit," refers to types of payment that may be included in the income of the owner of any kind of productive factor. These types of payment are distinguished in terms of (1) the minimum payment necessary to induce a factor into a specific use, (2) the payment a factor owner expects to receive when the factor enters or remains in a specific use, and (3) the actual payment received. Rent is the excess of a factor's expected return or income over the minimum necessary to bring forth the particular service to its specific use. Cost is the minimum necessary payment, while profit is the excess of the actual payment received over the expected payment when the resource was committed to the specific use. The term "entrepreneurial rent" is frequently used to describe the expected return to a firm over and above the minimum necessary to induce the firm to enter or remain in a specific industry. It corresponds to what is sometimes loosely thought of as profit.

That part of a factor payment termed rent by the present-day economist is, like the classical economist's rent, price-determined rather than price-determining. It is important to realize, however, that the distinction between cost and rent is a flexible one and depends critically upon the range of alternatives considered and upon the time period allowed for adjustment to changed conditions. For the economy as a whole, the return to many human and material productive services is wholly rent, since no other uses are available for and hence no opportunities are forgone by some employment. To any specific user of a particular productive service, in contrast, the payment he makes is wholly a cost if the productive service could earn as much from another firm in the same industry. In general, the broader the range of alternatives open to a productive service, the larger is the portion of its income that is a cost and the smaller its rent.

The distinction between cost and rent also depends upon the time period for adjustment one considers. For very short time periods most human and material assets are in effect fixed in their current uses. Thus, any return to their owners over and above the minimum return necessary to insure



that they be employed at all is rent. The longer the period of adjustment one considers, the greater the number of alternative employments that become available and the smaller the rents. In the longest of all long runs many factors of production have a great number of alternative employments, and in consequence their incomes are mostly cost. The term "quasi rent," introduced by Alfred Marshall, is used for a payment, especially to the specific material embodiment of nonhuman capital, which is rent in the short run but a cost in the long run, that is, when the capital is no longer fixed in a specific material form.

Finally, it should be stressed again that in the modern sense of the term, rent may be a part of the income of any productive factor. It has already been pointed out that the incomes accruing to material assets are largely rent so long as the specific material form of capital is fixed. In addition, actual wage and salary payments consist in part of rents in differing degree, depending upon circumstances already noted. The salaries of college professors, for example, may be largely a cost to any particular institution that employs them. From the viewpoint of all institutions of higher learning, however, the excess of the incomes of professors when adjusted for nonmonetary advantages over the incomes they might earn in government or industry is rent. Considering all potential employers of college professors, most of a professor's income is rent once he has made his occupational commitment. The same expected income may be largely a cost, however, before an individual chooses to enter college teaching in preference to some other occupation.

### Determination of rent

In this section the principles of the determination of rent are discussed. Rent as rental, or hire, value is considered first; this is followed by a discussion of the determinants of rent as an excess of expected return over cost. Because rent is so often thought of as the return to land, special attention will be given to the level of land rentals generally. This section will conclude with a discussion of the share of returns to land in the national income. The discussion of the returns to land is continued in the third section, where differences in land rentals attributable to differences in location are discussed.

**As rental value.** The determination of the rental, or hire, value of an asset is but a specific instance of the determination of the prices of productive services generally. In any specific use or industry

the rental value of a particular material asset is equal to the price of the final product multiplied by the marginal physical product of the services of the asset in the specific use, provided that competition prevails in both the product and the factor markets. The marginal physical productivity of the asset's services, in turn, varies directly with its scarcity relative to other productive factors. It is analytically useful, as in any problem of price determination, to separate the forces affecting rental value into (1) demand for the services of an asset in a particular industry as derived from the demand for the final product produced by the industry, (2) the supply of other productive factors to the industry, and (3) conditions of technology. For the economy as a whole, the demand for an asset's services is simply the summation of all the individual industry demand curves, provided that conditions of supply and technology for the economy as a whole are used in deriving the industry factor demand curves to be summed.

An increase in the demand for final product tends to raise both the rental per unit of an asset and the aggregate of rental payments to the asset in question. In general, an increase in the supply of other productive factors will tend to reduce the demand for the factor in question as these other factors are substituted for it in production. (It is possible, of course, for an increase in the supply of complementary factors to increase the demand for the asset in question, but on balance all other factors taken together are substitutes.) On the other hand, the increase in supply of other factors leads to an expansion of the industry's output and, hence, an increase in the demand for the asset in question. The strength of the increase in demand operating through the expansion of output is greater, the greater the elasticity of demand for the industry's product. The substitution effect varies directly with what is technically known as the *elasticity of substitution in production* between the asset in question and all other factors, that is, the relative change in the ratio of the input of the asset in question to all other factor inputs, divided by the relative change in the inverse ratio of their marginal physical productivities when output is held constant. Its value varies from zero, in the case where the asset in question is used in fixed proportions with all other factors, to plus infinity, where the asset in question and all other factors are perfect substitutes. The net effect of an increase in the supply of all other factors will be to increase or decrease demand for, and hence rental value of, the asset in question, depending upon

whether the elasticity of industry demand is numerically larger or smaller than the elasticity of substitution in production.

A technological change that increases or reduces the marginal productivity of the asset in question but leaves output unchanged will increase or reduce its demand and rental value. A neutral technological change, or one that increases the marginal productivity of the asset in question and all others in equal relative amounts, will increase or reduce the demand for, and rental value of, an asset, depending upon whether the elasticity of demand for the industry's product is numerically greater than or less than one.

In applying the above propositions to land rents it is important to distinguish between agricultural and urban uses of land. It would appear that the demand for agricultural output is highly inelastic with respect to its relative price and that the elasticity of substitution is about unity. On the other hand, the price elasticity of demand for housing, by far the most important use of urban land, is probably unity or even larger numerically, and the demand elasticities for many other urban products may well exceed unity. Since aggregate land values would appear to be large relative to the total value of structures near the centers of cities, where rentals per unit of land are high, the elasticity of substitution of land for other factors in producing most urban products is probably less than unity. Thus, it would appear that an increase in the supply of nonland factors would reduce the demand for, and rental value of, agricultural land and increase them for urban land. Also, a neutral technological improvement would reduce the rental value of agricultural land but leave unchanged or increase urban land rentals.

The supply schedule of a particular productive service to a specific use or industry is merely the aggregate supply schedule of the factor less the demand schedule for the factor in all other uses. At the edge of cities, for example, the total quantity of all land is fixed, and the supply of land for residential and other urban uses is this fixed quantity less the demand for agricultural land. The demand for the output of agricultural firms in the vicinity of cities is highly elastic, since this output is but a small part of the total national or world output. One would therefore expect the agricultural demand for, and the urban supply of, land at the edges of cities to be highly elastic.

Little can be said generally about the aggregate supply schedule of a productive factor facing the economy as a whole. However, the shape of the

aggregate supply schedule depends critically upon the length of time allowed for adjustment, certainly so for nonhuman agents of production. Probably because it is quite costly to add rapidly to the stocks of most material assets, their supplies tend to be highly inelastic over short periods of time. In the long run, the relative supply of most material assets—for example, houses as opposed to office buildings—may well be highly elastic. Since it would appear that the rate of saving is relatively insensitive to the rate of interest, the aggregate supply of all material assets or of capital to the economy is likewise probably relatively insensitive to their rental value.

An increase in the supply of an asset will lower its rental value per unit, but the aggregate rental value of all assets of a given type will increase or decrease depending upon whether the price elasticity of the factor demand schedule is numerically larger or smaller than unity. Provided that the supply of other factors to a specific industry is highly elastic, the demand elasticity for the services of a particular asset is a weighted average of the (negative of the) elasticity of substitution in production and the demand elasticity for the product; the weights are respectively the fraction of the industry's receipts paid out to all other factors and to owners of the asset in question. Thus, the demand for the services of land is likely to be inelastic—in agriculture because of the inelastic demand for the final product and in housing and other urban uses because of the less than unit elasticity of substitution in production. In either case, an increase in the supply of land, as might be brought about, in effect, by an improvement in transportation, would reduce the aggregate rental value of land.

As excess of expected return over cost. All of the forces that affect rent in the sense of rental, or hire, value also affect rent as the excess of expected return over cost. For any particular unit of the productive service actually devoted to the specific use, rent is merely the difference between the expected market price and the ordinate of the factor supply schedule at that quantity at which the particular unit enters into the specific use. In Figure 1, for example, units of the productive service that would enter the specific use at a price per unit of  $OD'$  receive a rent equal to  $D'E$  when the market price is  $OE$ , as under the conditions illustrated. The aggregate rent paid out to all units of the factor employed in the specific use is the integral of expected price minus the ordinate of the supply schedule taken over all units of the service employed in



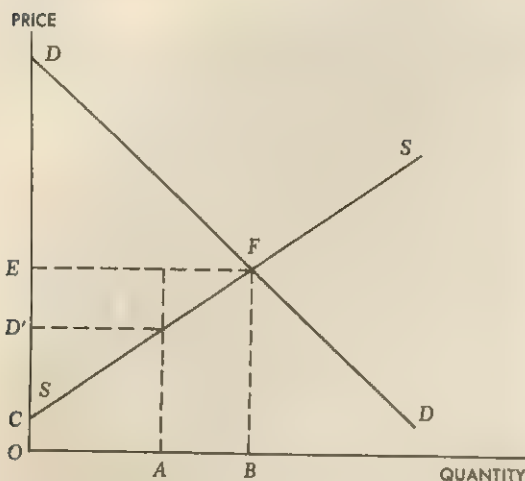


Figure 1

the specific use, the triangle *CEF* in Figure 1. Clearly, any event that causes the demand for the factor to increase, increases the rent received by each unit of the productive service and the aggregate rent received by all taken together. An increase in the supply schedule reduces the rent received by any given unit of the productive service. However, an increase in supply may either increase or decrease aggregate rent paid. If, for example, the supply schedule is perfectly inelastic, the whole of the expected payment to the factor is rent, and this total payment increases or decreases as demand is elastic or inelastic.

Rent as an excess of expected return over cost depends critically upon the slope of the supply schedule. Given the point of intersection of the demand and supply schedule, the steeper the supply schedule, the greater the rent received by any unit of the service and the aggregate rent received by all. In the case of labor, or the services of human beings, the slope of the supply schedule for a particular occupation varies directly with the extent of differences among individuals in their abilities to perform specific functions and, perhaps more importantly, with differences in their evaluations of the nonmonetary disadvantages or attractiveness of the occupation relative to others. Nonmonetary considerations are generally of little importance for most nonhuman assets.

**Share in national income.** The final topic considered in this section is the share of aggregate land rentals in the national income. This share, or the share of any factor, depends critically upon the elasticity of substitution between it and all other factors in production. An increase in the

supply of other factors relative to land increases the aggregate of all land rentals relative to the national income, depending upon whether the elasticity of substitution is less than or greater than unity. A neutral technological change likewise increases land's share if the elasticity of substitution is less than unity, provided that, as is probably the case, the elasticity of supply of land to the economy as a whole is less than that of all other factors. Of course, a technological change that increases land's marginal productivity and reduces that of other factors, leaving output unchanged, will increase land's share.

It has been asserted earlier that in agricultural uses the elasticity of substitution of land for other factors is about unity, while for urban uses of land this elasticity would appear to be less than unity. For the economy as a whole, the elasticity of substitution also depends upon the substitutability of agricultural for urban products in consumption. The very low relative price elasticity of demand for farm products suggests that substitution possibilities in consumption are small. This last consideration, coupled with the fact that urban output is large relative to farm output, suggests that for the economy as a whole the elasticity of substitution of land for other factors is probably less than unity. While data are very scanty, it is probably true that the returns to land are a smaller fraction of the national income in the more highly developed economies, including the United States, and have grown less rapidly over time than has national income. Such tendencies might result from the relative increase in land supply brought about by improvements in transportation with economic development or by the fact that the demand for land-intensive commodities, such as farm products, grows relatively less rapidly than the national income.

**Location rents.** This section continues the examination of the rental value of land by considering differential rentals attributable to differences in location. Such differentials are frequently referred to as "location rents" and are the major source of differences in the rental value of different parcels of urban land. While of interest for their own sake, differential land rentals are especially important because of their relation to the spatial organization of economic activity.

The first economist to undertake an investigation of the spatial aspects of economic activity in any detail was von Thünen. He postulated a single city located in a boundless plain of land of homogeneous quality in which transportation costs varied with distance from the city but were otherwise invariant. (This is what has since come to be

known as a "transportation surface.") He then showed that different types of farming would be arranged in annular (that is, ringlike) zones around the city. Commodities for which transport costs were relatively high—milk, for example, because of its perishability—or for which land is relatively unimportant would tend to be produced in the inner annuli, and commodities for which transport costs were low and land important would tend to be produced in the outer ones. Land rentals declined from annulus to annulus, and within a given annular zone the difference in land rental was equal to the difference in costs of production plus transportation to the market in the city.

In most developed countries today agricultural produce tends to be traded in national or world markets. Transportation costs to a particular market, except for a few commodities such as dairy products, have relatively little effect on the location of agricultural production. Thünen's analysis, however, is highly relevant for the analysis of land use in cities or urbanized areas and, in fact, is the basis for most modern economic theories of city structure. It also bears a striking similarity to the concentric-zone theory of city structure, developed by the sociologist Ernest W. Burgess about a century later. In theories of city structure, the central business district (CBD) plays the role of Thünen's isolated city. Located near the hub of the city's internal transport system, the CBD is generally the area of maximum accessibility to the local market for commodities and services of many kinds and to the local labor market. In addition, since rail terminals are typically located adjacent to the CBD, it is close to outside markets and sources of supply. (With the advent of highway and air transport, however, the comparative advantage of the CBD on this score has declined.) Since the CBD is generally the largest single employment and shopping center in the city, accessibility to the CBD is especially important for residential users of land. [See the *biographies* of BURGESS; THÜNEN.]

In the analysis of the relation of land rentals to the location of economic activity in cities, four distinct kinds of considerations are relevant. First, provided that competition prevails, any given user of land takes the prevailing level of rentals in any location as given. The use of land relative to other productive factors is governed by the principle that to minimize the cost of producing any given output, the firm adjusts its inputs in such a way that the marginal physical product per dollar spent is the same for all inputs. Thus, where land rentals are high relative to the prices of other productive services the marginal physical product of land is

relatively high. This means, for example, that high-rise apartment buildings tend to be built where residential land rentals are highest, as do loft-type or multistoried factory buildings in commercial areas of relatively high land rentals. Conversely, single-family detached houses and single-story factory buildings tend to predominate in areas of relatively low land values.

Secondly, for firms of any given industry to be in locational equilibrium, it is necessary for land rentals to vary directly with the price received for their product and inversely with the price paid for productive factors other than land, both being measured at the point of production. If, for example, firms located closer to the CBD received higher prices for their products but all firms paid the same rentals per unit of land, firms closer to the CBD would earn larger incomes. It would then be in the interest of firms located further from the CBD to offer more for the closer sites than firms located there were currently paying. In the process, the rental of sites closer to the CBD would be bid up, while that of more distant sites would be bid down. In general, the conditions of firm equilibrium require that the relative variation in the rentals of land used by firms in a given industry be greater, the more product or nonland input prices vary relatively with respect to location and the smaller the share of land in the receipts of the industry. The variation of final product and input prices with location results primarily from transport costs. Transport costs on final products or raw materials, in turn, tend to be greatest for small shipments and shippers, for bulky and difficult-to-handle goods, and for goods of high value in proportion to weight. The costs of transporting people, both as buyers of final products and sellers of labor, are also important determinants of rental gradients.

The third consideration relevant in the determination of differential land rentals might be called equilibrium in the market for land. Equilibrium requires that each parcel of land be devoted to that use which yields the highest rental. For if this were not the case, landowners could increase their incomes by leasing their land to firms in alternate industries. Both the amount of rental offered for land at any location and the rate of change of rental with change in location, or rental gradient, are important in allocating land among industries. If in the vicinity of any center of activity, or local peak in land rentals, firms of two different industries are to locate, firms of the industry with the steeper rental gradient will locate closer to the center. This follows from the fact that if the func-



tion showing the rental one industry offers for land is anywhere above that for the other industry, it will be so at the center. Therefore, as one proceeds out from any local peak in land values, the relative decline in land rentals will diminish in passing from the area of location of one industry to another.

It should be noted that taken together, and in the absence of external economies in land use, the conditions of locational and land-market equilibrium imply that the rental value of land is maximized. Maximization of rental value is requisite to efficient resource use. In actual practice, of course, external economies in urban land use may be of great substantive importance, and in the absence of conscious social control the bidding for land by private individuals and business firms may fail to maximize the rental value of land. This failure provides the economic rationale for planning, zoning, and other forms of governmental land-use control. Now, it is frequently stated that the aim of governmental land-use control is the minimization of land rentals. The reasons for this view are interesting but not relevant here. Its absurdity is readily seen by considering that an absolute minimum of land rentals could be achieved by forbidding land to be used for any purpose whatsoever.

Finally, for any pattern of locational differences in land rentals to be an equilibrium pattern, it is necessary that there be no unsatisfied buyers or sellers of any commodity at any point in space. For if there were, the prices of some commodities would rise or fall, and the maximum rentals offered for certain parcels of land would likewise change. As a result, it would be profitable for some landowners to change the use to which the land they own is put.

Having outlined in very general terms the principles underlying the determinants of differentials in land rentals with location, let us consider some specific forces governing actual rental differentials. It was noted earlier that the CBD is the area of maximum accessibility or minimum transport costs for the local product and labor market and that it is convenient to the rail lines serving as links to outside markets and material-supply sources. The relatively high cost of local movement because of traffic congestion in the center, especially during rush hours, may mean that prices received for products rapidly fall off or wages paid increase rapidly with movement away from the center. Where actual transport of goods is involved, the products of CBD firms are generally of high value in relation to weight and often involve small shipments or shippers, so that transport costs are relatively high. Finally, for the commercial, trade, and

service firms that locate in the CBD it would seem that land is a relatively unimportant factor of production.

Agriculture, on the other hand, is probably the most land intensive of all economic activities, and shipments of agricultural products are frequently large and involve products of relatively low value in relation to weight, so that transport costs are relatively low. Thus, commercial and retail firms, residences, and agricultural firms are located in roughly annular zones surrounding the CBD, and land rentals decline at progressively lower rates in each of these zones.

There are, of course, many exceptions to the broadly annular zonal pattern of location and corresponding relative declines in land rentals surrounding the CBD. The two most important are the concentration of manufacturing in the vicinity of waterways, rail lines, and truck routes and the more or less regular hierarchical pattern of retail and service business. It is well known that the structure of transport costs is such that manufacturers and other processors tend to find locations at point material sources or markets less costly than intermediate ones. While sources of raw materials are rarely found inside city limits, rail lines, freight and truck terminals, and waterways are essentially point or linear material sources and markets for final products for shipment outside the city. The so-called light manufacturers, firms receiving or sending shipments in less than carload lots, find location close to freight terminals desirable. The same is true for wholesalers whose outbound shipments are frequently less than carload lots. Since land is probably a more important productive factor for light manufacturers and wholesalers than for commercial, retail, and service firms in the CBD proper, the former group tends to locate around the outer edges of the CBD. In this area of location, land rentals probably decline relatively less rapidly than within the CBD proper, although more rapidly than in the surrounding residential zone. Heavy manufacturers or firms receiving or making shipments in carload lots would find locations anywhere along rail lines or truck routes equally attractive. As a result there is no reason relating to transport costs on commodities to expect land rentals in immediate proximity to such routes to vary with location along them.

The tendency for retail and service business to concentrate in a hierarchy of centers of various sizes at more or less regularly spaced intervals or along major streets within a city is explained by central place theory [see SPATIAL ECONOMICS]. In addition, because certain facilities, such as parking lots, are shared, some costs are lower in shopping

centers than elsewhere. The demand for any particular firm's product is probably greater within a center because of the tendency for customers to make a variety of purchases during any given trip and to comparison shop, and the vicinities of certain points, such as intersections of major streets and rapid transit stops, are points of local minimum transport costs for many kinds of firms. The reasons given for the relative steepness of rental gradient within the CBD apply, but with diminished force, in these non-CBD retail and service centers. Hence, the rental value of land is likely to decline relatively rapidly in the vicinity of these outlying centers. It would appear that the peaks of land rentals are relatively higher in the centers belonging to higher levels of the hierarchy, the highest of which is the CBD, but there is no reason to expect the peaks of rental value in any given level of the hierarchy to vary systematically with location in the city.

Within the residential zone housing prices decline with distance from the CBD. If this were not the case, households located closer to the CBD, and bearing lower transport costs to and from it, would be better off than households in more distant locations. It would then be in the interest of the more distant households to offer more for housing closer to the CBD than its current inhabitants paid, and a decline in housing prices with distance from the CBD would result. Now, it can be shown that a necessary condition for household equilibrium is that the relative rate of decline in housing prices per mile, say, be numerically equal to a household's additional expenditures on transportation per mile divided by its expenditure on housing. Households with relatively high transport costs, such as those with more than one worker employed in the CBD, have an incentive to locate close to the CBD. Those who spend greater amounts on housing likewise have an incentive to locate at greater distances. Thus, the relative rate of decline of housing prices is smaller in the outer parts of the city, and for this reason land rentals would decline at relatively slower rates at greater distances from the center. This last tendency is at least partly offset, however, by the fact that because of the less than unit elasticity of substitution of land for other factors, the relative importance of land in producing housing is smaller in the outer parts of cities. Because transport costs tend to be smaller for households located in the vicinity of rapid transit routes and express highways, land rentals decline relatively less rapidly with distance from the CBD in the neighborhood of these facilities.

Housing prices in cities, and thus land rentals, may vary for many other reasons. Since accessi-

bility to local shopping and employment centers, rapid transit stops, and educational, cultural, or recreational centers has value because of the saving in transport costs, one might expect local variations in residential land rentals in the vicinity of the centers. In addition, the prices people are willing to pay for residences are undoubtedly influenced greatly by the character of the surrounding area. It would appear, for example, that housing prices, and thus residential land rentals, are lower in the immediate vicinity of local manufacturing centers and higher in higher-income neighborhoods and in the immediate vicinity of attractive natural surroundings.

RICHARD F. MUTH

[See also SPATIAL ECONOMICS.]

#### BIBLIOGRAPHY

- HICKS, JOHN R. (1932) 1964 *The Theory of Wages*. New York: St. Martins.
- HOOVER, EDGAR M.; and VERNON, RAYMOND 1959 *Anatomy of a Metropolis: The Changing Distribution of People and Jobs Within the New York Metropolitan Region*. Cambridge, Mass.: Harvard Univ. Press. → A paperback edition was published in 1962 by Doubleday.
- HOYT, HOMER 1933 *One Hundred Years of Land Values in Chicago: The Relationship of the Growth of Chicago to the Rise in Its Land Values, 1830-1933*. Univ. of Chicago Press.
- MARSHALL, ALFRED (1890) 1961 *Principles of Economics*. 2 vols. 9th ed. New York and London: Macmillan. → See especially Book 5, Chapters 9-11, and Book 6, Chapter 9.
- MUTH, RICHARD F. 1961 *Economic Change and Rural-Urban Land Conversions*. *Econometrica* 29:1-23.
- RICARDO, DAVID (1817) 1962 *Principles of Political Economy and Taxation*. London: Dent; New York: Dutton. → A paperback edition was published in 1963 by Irwin. See especially Chapter 2.

## REPRESENTATION

- I. THEORY
- II. REPRESENTATIONAL SYSTEMS
- III. REPRESENTATIONAL BEHAVIOR

Alfred de Grazia  
Giovanni Sartori  
Kenneth Janda

#### THEORY

Representation may be defined most usefully as a relation between two persons, the representative and the represented or constituent, with the representative holding the authority to perform various actions that incorporate the agreement of the represented. The relation is by no means simple, since practically every type of human communication and perception can be shown to be intrinsic to representation. The relation is sociopsychological. Essentially subjective, it may, however, be affected by numerous objective conditions and events.



Representation is a concept of social interest largely in the contexts of power relations among leaders (representatives) and followers (constituents)—whether in government, church, school, business, or the family. It may well be noted in terms of a scale ranging from nonrelation (or “bad” relation) to perfect or full relation. Thus an accord of  $x$  degree has to be achieved before representation may be said to exist; below that point, A is said not to be represented by M.

### The study of representation

The simplest index of representation is a positive answer that a person may give to the query: “Do you feel represented by M?”

What this degree of accord is to be is not standardized, if only because the very parameters of the relation are unexplored. When H. F. Gosnell and this writer attempted such an exploration in 1940 (see Gosnell 1948), it was plain that the classical utterances on political representation were commonly polemical, unsystematic, and superficial (cf. Burke 1774; Mill 1861; the materials presented in Loewenstein 1922; Clarke 1936; Smith 1940). As soon as the thicket of representative relations was entered, progress became slow. Obviously representation must consist of unconscious as well as conscious relations, expressive as well as sanctioned actions, unknown as well as known actions. Otherwise one falls into the trap of regarding as determinative of the representative relation solely conscious, known legislative behavior—parameters that are more often missing than present in the setting about which the scholar wishes to say “A is represented by M.” This rationalistic fallacy has prevented generations of scholars of democracy from coming to grips with problems of representation (de Grazia 1951).

Today, by contrast, the study goes forward. John Wahlke, Heinz Eulau, William Buchanan, and L. C. Ferguson (1962) managed to perform the difficult task of fitting standardized inquiries of several sets of U.S. state legislators into the real-life environment of representation and attaching to the representatives’ cognitive and perceptive structures and functions a set of postures toward representation: the *expert* role defines the ability and reputation of a member with respect to the work of the assembly; the *purposive* role includes the ritualist and the tribune who defends the people; in the *representative* role some legislators take up a trustee posture toward their constituents, others take a delegate posture, and some “politicos” are disposed to both in various ways; finally, the *areal* role denotes the tendency toward identifying with

the elective district, on the one hand, and the whole state, on the other. At about the same time, Warren E. Miller and Donald E. Stokes (1963) undertook to question samples of the electorate of various districts concerning their congressmen.

As they have exposed the facts of representation, recent studies have also disclosed the materials upon which the long-argued ethics of representation must rest. The question “Whom should a representative represent?” can be answered precisely, granted always that one knows what it is he wishes the representative system to produce by way of action or expression.

**Clarifying the concept.** Despite the improvement in research, the connection of representative and constituent in any setting or in society as a whole must continue to be referred to imprecisely in absolute terms (“is”—or “is not”—“represented”), or in imprecise comparative terms (“more representative than”). What would be required to make the terminology more precise?

First of all, a clear, scientific approach to the myth-shrouded concept is required. Considering the wealth of empirical investigation in human relations, it is odd and regrettable that so little work has borne directly upon the foremost structural problem of democracy.

The state of knowledge of representation would also be improved if translation of concepts and findings of other scientific fields were made. Since representation is a psychological condition, there are obviously numerous relevant concepts, such as perception (awareness), cognition (information), communication, stereotypes, and authority (charisma), that are to be regularly taken into account.

In addition, there are many group-behavior concepts that pertain to representation. If representation is a relation between A and M, it is also a relation between B, C, D, . . . , on the one hand, and M, on the other, the group boundary of the members being the electoral constituency, which in turn can take numerous territorial or functional forms and can exclude a few or most of the population.

**Determining representativeness.** What determines the representation of a group’s members and the selection of their formal leaders? Numerous variables do. They may be either constructed or naturally occurring variables, such as the social composition of the group and its aggregate attitudes toward personality types and public issues. In all attempts to influence or study these variables, it must be remembered that every force brought to bear upon A will perhaps affect B, C . . . , differently. This obvious fact has highly complicated

meanings for electoral study. The effects of propaganda, for example, although difficult to measure, are highly germane to the study of representation. A major source of change in representation may be the propagandistic operations of parties contending for the right to be representatives.

If propaganda is regarded as stimulus, intended to cause responses favorable to the propagandists on the part of the propagandized (admittedly the represented are continuously being propagandized), the chances of success of the stimulus rest upon its penetration of the more basic norms of the population. If, for example, the expectations and demands of a group are heightened by the winds of an ideology new to its members, or by re-education according to nontraditional principles, increased pressures will be exerted upon those who offer themselves as representatives. The latter will be obliged to respond by messages of a confining or, more likely, exhorting nature.

In sum, the study of the representation of a group results from the special algebra of aggregations of individuals. Two theories of representation, one for individuals and the other for groups, are not necessary. Admittedly, if it is difficult to assess the condition of representation of a single person, extending a set of such single calculi to cover an aggregate of thousands would give findings of dubious validity. Still, counterbalancing and averaging errors of individual estimates permit the result that a statement made about the representation of a group is likely to be no less precise and may even be more valid than one made about an individual.

### Representative government

Thus far, we have spoken of individual and group psychology in representation as being affected by subjective and symbolic determinants. There also exist as determinants a great number of institutional devices to produce (or restrict) representation. Election systems, nomination systems, campaigning systems, etc. are the devices. These devices, which may sometimes be called institutions but often are of a relatively minute nature (such as a term of office), are to be construed as recurring sets of similar actions relevant to the relation of representation—habitual, routine, predictable, and enforced by specialized personnel. There is no need, therefore, to separate the study of representative government from the study of representation.

A group of institutional devices intended to produce a high level of representation in a population may be called representative government. Historically, many governments have been representative

without being representative governments, since the conditions for representation may be "ideal" without the employment of institutional devices. One can readily point to direct democracies, which have existed in many times and places—the ancient Greek cities, areas of sub-Saharan Africa, the Italian cities of ancient, medieval, and early modern times, the Teutonic tribes, and the New England town meetings. They need not be defended as models of representation in order to accept the premise that their peoples received on the average as much representation as, or more than, other peoples have received under representative governments. Actually, were it not for the fact that a largely mythical distinction is made between "true" representative government and "direct democracy," we should lump the two together, not only because of the similar effects produced, but also because of principles and machinery of choice (elections) and control employed in both cases: the Athenian lottery brought forth officers who were supposed to represent the citizens in the same sense as the presidents chosen by modern election.

More bothersome to theory is the absence of the devices of representative government in certain groups where a high degree of over-all representation exists, such as Nazi Germany—between 1935 and 1942—Augustan Rome, and Napoleonic France, and, worse, the apparent absence of representation in certain representative governments, such as the United States under John Adams and the Soviet Union under Stalin.

**Institutional devices.** What is the nature of this cluster of devices—nomination, apportionment, election, ballots, campaigns, etc.—that is called representative government? Theoretically they have to be regarded as procedural constructions intended to form representation in a desired way. Thus, an apportionment that follows local cultural boundaries is expected to produce representatives colored by local ideas, and one that cuts across boundaries in a nearly random way is supposed to suppress established localism. A single-member district system in which a majority or plurality vote determines the winning candidate is expected to accentuate the majority sentiment of each district; whereas a multimember district in which any considerable minority is bound to elect at least one candidate, is a device intended to enrich legislative discussion by presenting opposing views from the same constituency.

Thousands of such variant election systems might exist, although, because of imitative forces, there are in fact mere hundreds. Only the fool or ideologue insists that one system is the only system



of representative government. Indeed, as we have indicated, it is impossible to show that representative government produces a strikingly larger representation of the population than other forms of government. However, as representative government becomes more understood, and more controllable, its devices may be arranged to produce with fair accuracy the fuller representation of whomever the system is set up to represent.

It is important to stress that the devices of representative government are not all things to all people. They help certain people to get more representation, and this is performed almost always at the expense of someone else's representation. Each successive change in history (and in the great many short-term cycles of history that occur from year to year in contemporary society) has meant a redefinition of who is "worth" representing in a people. Attitude, procedure, and device always go together, one often canceling out the effect of another. For instance, the kind of representation available to citizens of American states from their state legislatures was changed by court order in *Baker v. Carr* (1962) and *Reynolds v. Sims* (1964) to provide districts of equal population, but simultaneously the total representativeness of the legislatures was, perhaps temporarily, reduced when many unfamiliar candidates presented themselves to the people. In a more permanent way, perhaps, the political power of rural and small city areas was cut down. (For other effects of these cases on the condition of representation, see de Grazia 1963.) Disturbing a set of relations that is premised on awareness and acquaintanceship can easily reduce representation in the very setting where new devices are intended to promote it.

Thus the devices of Italian and German representation were modified after World War I to permit more accurate manifestations of public opinion in the legislatures. Proportional representation was employed, for example. Some of the effects were as foreseen—representativeness was increased—but the largest effect of the innovations was to promote a contrary form of representation, based upon the *Duce* and *Führer* principles, i.e., direct, controlled representation of the people, emphasizing uncontrolled, unrepresentative activity on behalf of the mass (Hermens 1941). In quite another way, far from "strengthening the state legislatures," as the Supreme Court maintained in associated opinions, the aforementioned 1962 and 1964 American cases could perhaps hasten the decline of federalism and, within the states, diminish the role of the legislature in relation to that of the governor.

These instances permit a proposition about changes in representation. Representation is broader

than the devices that regulate it. Classical speculation and debate over representation have almost always looked on representation as a linear scale relation: a device would either increase or diminish representation. This is true, or can be made to be true, if the question that operationally defines the relation is simplified; for example, one may ask, "Do you feel more or less represented today than five years ago?" Alternatively, representation can be confined, usually unconsciously, to a certain kind of issue: "Has your representative followed your wishes with respect to voting on the budget?" A man with an atomistic view of society may consider this an adequate test of whether representation exists. However, many changes in representation could occur that would not be tapped by such questions and therefore would not be measured at all and might lead to a false general interpretation.

One must conclude that the burden of plotting and mapping the effects of disturbances and changes in most devices of representation is too heavy for contemporary political science to bear and far too heavy for the much more limited rationality of practical politics. Although each institutional device can be perceived to carry certain effects, such effects are not well known, either scientifically or in applied politics. If its individual effects can be predicted, its indirect effects when taken as part of a system of devices cannot; the coming of other changes cannot be foreseen either, and these may be of a character so gross as to overwhelm the effects of the initial change.

### The role of representation

Representation is multifaceted. It is a broad relation. To define it otherwise invites continual frustration. To be useful in the traditional scope of its usage, it has to be defined as broadly as the range of expectations that people possess in regard to government, and hence in regard to public officials and leaders (*pari passu*, representation within all social organizations should be so defined).

Since many traits, events, and institutional devices affect the state of representation, the represented is approachable along numerous paths. It may, therefore, be sensed as only a small loss to him to be deprived of a vaguely familiar, "trivially" occupied district representative, when at the same time he is receiving a new charge of emotional satisfaction from a physically remote but psychologically near national figure. It is as useless to berate him for shortsightedness, romanticism, and foolishness as to berate any other mystic who shuns the practical. Indeed, this mysticism is more practical, for it costs nothing in self-discipline.

One conclusion about this particular phenomenon may be that only a self-disciplined ruling class, inured to the presences of commanding personalities and individuals of high status, can operate a republic over a long period of time; the mass of the population must succumb sooner or later to the seductive choice of the great over the petty, majesty over meanness, feeling over calculation. This elite need not be one of birth—it can emerge from the great population—but it must have two traits if it is to function in the bosom of modern mass democracy: (a) it must have a means of maintaining solidarity on the principles of multifaceted and rationalistic representation despite the surge of the mass of people toward single-executive leadership, and (b) it must have the skills to govern the masses.

The crisis of representation in the twentieth century has been precisely the inability of the old ruling class that created the rich fabric of representation to maintain its solidarity and guide the whole population. Thus in Russia, Italy, Germany, Spain, Japan, Argentina, and France, to name only the important catastrophic cases, the exaltation of executive representation overcame the intricate and ramified system that had developed.

Today, after the historical experience provided by these cases, public opinion seems in a less dangerous state. Yet the United States and the Soviet Union both face permanent crisis over the mode of representation to be provided their populations. The Soviet Union depends on an ideological elite party to ward off single-executive dictatorship. The United States depends on a complicated legalistic set of representative devices and localized parties and elites. Neither is safe; neither is advancing toward the vision of representation developed according to humanistic principles whereby, as John Dewey described it, a citizen lives a life of enriching communion with his fellow men in an associational society (1927; cf. Follett 1918).

The vision of a pluralistic or "mobilistic" system that represents plural types of men and plurally complicated persons stands in sharp contradiction to the idea of monolithic, primitive, emotional representation via the central person, or "hero," of the society. It is this system that might produce the republican elite referred to above as the prerequisite to an enduring democracy. It is, furthermore, a modern concept, quite at home in a multigroup society where "human systems engineering" has become a vigorous part of industrial technology. How to implement it presents many a problem now far beyond the capacities of scientists to solve and politicians to apply. By contrast, the notion of a single-faceted representation by a hero is so pow-

erful in the human subconscious and so much easier to implement that throughout history it has achieved a far greater popularity.

ALFRED DE GRAZIA

[See also DEMOCRACY; ELECTIONS; LEGISLATION, article on NATURE AND FUNCTIONS; MAJORITY RULE; RESPONSIBILITY. Other relevant material may be found in the biographies of BURKE and MILL.]

## BIBLIOGRAPHY

- BRADY, ROBERT A. 1943 *Business as a System of Power*. New York: Columbia Univ. Press.
- BURKE, EDMUND (1774) 1920 *A Letter to the Sheriffs of Bristol*. Cambridge Univ. Press.
- CLARKE, MAUDE V. (1936) 1964 *Medieval Representation and Consent: A Study of Early Parliaments in England and Ireland With Special Reference to the Modus tenendi parliamentum*. New York: Russell.
- DE GRAZIA, ALFRED 1951 *Public and Republic: Political Representation in America*. New York: Knopf.
- DE GRAZIA, ALFRED 1963 *Apportionment and Representative Government*. New York: Praeger.
- DEWEY, JOHN (1927) 1957 *The Public and Its Problems*. Denver: Swallow.
- FOLLETT, MARY P. 1918 *The New State: Group Organization, the Solution of Popular Government*. New York: Longmans.
- GOSNELL, HAROLD F. 1948 *Democracy: The Threshold of Freedom*. New York: Ronald Press.
- HERMENS, FERDINAND A. 1941 *Democracy or Anarchy? A Study of Proportional Representation*. Univ. of Notre Dame (Ind.) Press.
- LEISENBERG, AVERY 1942 *Administrative Regulation: A Study in Representation of Interests*. Univ. of Chicago Press.
- LOEWENSTEIN, KARL 1922 *Volk und Parlament nach der Staatstheorie der französischen Nationalversammlung von 1789*. Munich: Drei Masken.
- MILL, JOHN STUART (1861) 1958 *Considerations on Representative Government*. New York: Liberal Arts Press.
- MILLER, WARREN E.; and STOKES, DONALD E. 1963 *Constituency Influence in Congress*. *American Political Science Review* 56:45-56.
- SMITH, THOMAS V. 1940 *The Legislative Way of Life*. Univ. of Chicago Press.
- WAHLKE, JOHN et al. 1962 *The Legislative System: Explorations in Legislative Behavior*. New York: Wiley.

## II

### REPRESENTATIONAL SYSTEMS

In order to define a representational system one must first define *representation*, a many-faceted and elusive concept. The term is associated with three quite different meanings: (1) the idea of *mandate*, or instructions; (2) the idea of *representativeness*, that is, resemblance and similarity; and (3) the idea of *responsibility*, or accountability.

The first meaning is derived from private law and belongs to the context of juristic representation, whereas the second meaning is derived from a sociological or existential context according to which representation is basically a fact of likeness



that transcends all voluntary selection (Friedrich 1963, p. 304) and even awareness (unconscious representation; Gosnell 1948, p. 141). In the juristic meaning, we often speak of a representative as a delegate or a mandatory; in the sociological sense we say that somebody is "representative of" to mean that he impersonates (*impersonare*) certain existential features of the group, class, or vocation from which he is drawn. The idea of responsibility and its relationship to representative government, i.e., the third meaning, will be examined below.

To be sure, we are concerned here only with political representation. But political representation cannot be dealt with in isolation from other kinds of representation. It is closely connected with sociological (or existential) representation on the one hand and with juristic representation on the other.

The extent to which political and sociological representation are intertwined can easily be detected in the discussions about electoral systems and *overrepresentation* or *underrepresentation*: a debate that expresses a concern for representativeness and resemblance. Yet the distinction between political and existential representation must be kept firm; otherwise almost *any* political system may claim to be somehow a representational system: ruling groups are always "representative of" the sections or the strata of the society from which they derive.

The extent to which political and juristic representation overlap is shown by the fact that the European theorists of representation—German, French, and Italian—have reached the almost unanimous conclusion that political representation is not "true representation," precisely by using the yardstick of a private law notion of representation (Laband 1876–1882; Orlando 1895; Saripolos 1899; Carré de Malberg 1922; Kelsen 1945). In fact, unless the distinction between political and juristic representation is kept firm, it is only too easy to reach the conclusion that *no* political system can claim to be a really representational system. Yet the distinction between political representation and juristic representation cannot be pushed too far because, among other reasons, political representation is formalized in our institutional arrangements and is part and parcel of constitutionalism.

### Historical development

The emergence of modern political representation from its medieval phase was a gradual process best seen in the political history of England during the second half of the eighteenth century and in the writings of Algernon Sidney, John Toland, and

Humphrey Mackworth. The French Revolution, however, provided a clear-cut break from medieval representation, a fact demonstrated in the constitution of 1791, which expressly forbade mandatory instructions to representatives because the "representatives nominated in the constituencies do not represent a particular constituency but the entire nation." Two subtleties are worth noting. First, by saying nominated *in*, the authors of the constitution expressly meant that the representatives were not nominated *by* their electors; second, the authors deliberately used "nation" instead of "people."

The distinction has far-reaching implications. If the *people* are declared sovereign, the will of the representatives of the people is secondary to the will of a principal (*dominus*), and two wills, that of the people and that of the assembly, are actually postulated. But if the *nation* is declared sovereign—as in article 3 of the Declaration of Rights of 1789—then there is only one will, for the will of the nation is the very same as the will of those who are entitled to speak for the nation (Carré de Malberg [1922] 1962, p. 267).

Thus the origin of the modern idea of political representation was vitally connected with the principle that the deputies represent the will of the nation, not the will of the people. Somewhat paradoxically, members of parliament took on the name "deputies" at the very moment that they ceased to be, properly speaking, deputies. Not only were they declared free agents who should not be given instructions, but they were called to represent a will that did not exist before their own will.

The idea of representation of the nation does not apply only to the origins of modern representation. The prohibition of mandates, sometimes explicitly connected with the principle of the sovereignty of the nation, is affirmed in the nineteenth-century and twentieth-century constitutions of most European countries: Belgium, 1831; Italy, 1848 and 1948; Prussia, 1850; Sweden, 1866; Austria, 1867, 1920, and 1945; Germany, 1871 and 1949; the Netherlands, 1887; and Denmark, 1915. Article 67 of the Italian constitution of 1948 is a good example: "each member of parliament represents the nation and performs his task without being bound by a mandate." And the tortuous formula adopted in the French constitutions of 1946 and 1958—"the national sovereignty belongs to the people"—makes little sense unless one is reminded of the declaration of 1789.

The pattern in the United States may appear to have been different in that no state constitution expressly forbids mandatory instructions. But this omission means only that American constitution framers did not feel the need to state the obvious,

since the United States did not have to break away from a medieval past. Likewise, the fact that the expression "sovereignty of the nation" has no legal meaning in English public law does not imply that British constitutional practice actually abides by a different presupposition. In effect, the French "will of the nation" entrusted to the assembly called for the same founding principle advocated by Burke in his Bristol address of 1774: that the representative should no longer be a delegate but a trustee acting according to his free judgment.

Modern representation reflects a fundamental historical change. Until the Glorious Revolution in England, the Declaration of Independence in the United States, and the French Revolution, representation was not associated with government. Representative bodies were intermediary channels between their mandatories and the sovereign: they represented *somebody* to *somebody else*. But the more the power of parliament grew and the more a parliament was located at the very center of the state, the more representative bodies took on a second function. In addition to representing the citizens, they also ruled over them. Present-day parliaments are Janus-faced: they represent the citizens to the state and the state to the citizens. Clearly, a parliament cannot acquire its modern function, that of governing, without change in its former function, that of representing. A representative body that is inserted into the state must be allowed the discretion that enables it to act for the state. Yet this transformation raises a number of thorny problems.

We are confronted with a puzzling fact: a point has been reached at which a representative body ends by representing somebody (the people, or the nation) to himself. The third party, the one to whom the second party was assumed to represent the first, has disappeared. Moreover, the principal of the representative relationship is hard to identify; for it is not easy to answer the question, *Who* is being represented? According to the constitutional prohibition of mandates and the emphasis on the representation of the *nation*, the representative does not, or should not, represent those who elect him. However, if a representative does not represent his electors, it follows that it is not the election that creates a representative. If this is so, the following question must be faced: Are elections a necessary means for the establishment of a representational system?

### Representation and elections

**Representation without elections.** The query may be stated as follows: Can there be representation without elections? A positive reply is often

given to this question on grounds that are irrelevant to the problem of political representation. If we refer, for instance, to existential or sociological representation, that is, to a mere *fact* of likeness, then it is clear that this kind of representation does not require an election. If representation is defined merely as an *idem sentire*, as the fact of "coinciding in opinion with" (Bagehot 1859), any method of selection or even no method at all might do. What matters is not the procedure that may best secure the coincidence in opinion (and behavior) between the representative and the represented but the existence of such coincidence. Political representation, however, is concerned precisely with securing this coincidence.

Aside from existential representation, one can still point to many instances in which a representative is appointed, not elected—an ambassador, for example. Yet even this case is hardly pertinent, since other guarantees besides the method of selection assure that he will perform his duty qua representative. By contrast, a member of parliament cannot be recalled at will and is subject to democratic control only because he may not be re-elected. The more political representation outgrows juristic representation, the less the former maintains the guarantees provided by the latter—except for the sanction of removal afforded by periodic elections. And this is why the method of selection appears all-important and becomes the peculiar concern of political representation. In principle there can be no representation (other than existential) unless those represented have some say and some way of protecting themselves; otherwise they would be at the mercy of their alleged representative. And since, ultimately, political representation provides only an electoral safeguard, *in this case* there can be no political representation without an election.

**Elections without representation.** The reverse is not true, for we can have elections without representation. Throughout history officials and rulers have been elected, but the election did not in the least imply that the elected represents his electors. This calls attention to the fact that representation is a "think so," in the sense that it is only in terms of "ideas" that a person can be "made present" by another person (Friedrich 1962, p. 164). With the sole exception of the marginal case of existential unconscious representation, there can be no representation unless the representative *feels* the expectation of those he represents, and feels that it is a *normative expectation*. Not only does representation imply "think so," but it is also inevitably linked with an "ought." Therefore, if an election does not explicitly carry with it the meaning and



the intention of electing a representative, not an absolute ruler, no representation follows from an electoral procedure. But this does not prove that elections are not a necessary means: it proves only that they are not, *per se*, a sufficient means.

**Elective representation.** Elections are one thing, and representation is another. Still, modern political representation is "elective representation," for it is precisely this combination that makes representation both *political* and *modern*.

The means (elections) cannot surrogate the animus (the idea of representation), but the animus by itself does not suffice. Nonelective representation—"virtual" representation in the vocabulary of Burke—needs the support and the safety of "actual" (electoral) representation.

Let it be emphasized that this conclusion applies *only* to political representation (not to juristic and even less to existential representation) and especially with reference to securing responsibility. The electoral theory of representation actually is the theory of responsible and responsive representation; it is not concerned with satisfying the requirement of similarity but with securing the obligation to respond. [See ELECTIONS.]

### Representational systems defined

**Conditions.** We may now attempt to reply to the question, What *is*, and what *is not*, a representational system? A survey of the literature indicates wide disagreement, for it would seem that *any one* of the following conditions is either sufficient or necessary:

(1) The people freely and periodically elect a body of representatives—the electoral theory of representation.

(2) The governors are accountable or responsible to the governed—the responsibility theory of representation.

(3) The governors are agents or delegates who carry out the instructions received from their electors—the mandate theory of representation.

(4) The people feel the same as the state—the *idem sentire*, or syntony, theory of representation.

(5) The people consent to the decisions of their governors—the consent theory of representation.

(6) The people share, in some significant way, in the making of relevant political decisions—the participation theory of representation.

(7) The governors are a representative sample of the governed—the resemblance, or the mirroring, theory of representation.

According to the previous discussion, conditions 1 and 2 are joint conditions. The mere occurrence of elections does not indicate, in itself, a repre-

sentational system; accountability to the people is a mere phrase if it cannot be enforced by the sanction of electoral removal. Condition 3 either refers to the premodern practice of representation or appears too exacting and impracticable. On the other hand, conditions 4, 5, and 6 are, *per se*, too vague. It is only too easy to presume consent and *idem sentire*; participation may well be obtained by manipulation and regimentation. It can be said, therefore, that these three conditions are more a consequence than a premise; they refer to something that is likely to be found *in* a representational system but not to a founding condition of a representational system. As for condition 7, the requirement of representativeness is an additional, not a necessary, requirement, which depends very much on the electoral system that is adopted.

**Responsibility and representativeness.** Nobody denies, to be sure, that one is likely to feel better represented if the representative is somebody "like himself," that is, like-minded, somebody who acts the same because he is (existentially, or vocationally) the same. While this explains the recurrent quest for a parliament that is related to the society (as Mirabeau said) like the map to the territory it represents (for example, Ross 1955, p. 51), the unfortunate thing is that one may conceive of a parliament that is a perfect mirror and yet does not perform a mirroring function. This is why accountability takes a procedural precedence over resemblance, at least in politics.

The argument can be pursued: even if representativeness is not, by itself, a sufficient condition, it can still remain a necessary requirement. The difficulty is, however, that while responsibility and like-mindedness are easily combined in a one-to-one relationship, they are very difficult to combine in a many-to-one relationship (when the many amount to tens and even hundreds of thousands). In the context of political representation we are, therefore, confronted with a dilemma: either we sacrifice responsibility to representativeness, or we sacrifice representativeness to responsibility. This conclusion needs to be qualified, however, and calls for a closer look into the notion of responsibility.

The idea of responsibility has two sides, or implications: (1) the *personal* responsibility to someone, that is, the obligation of the representative to answer to the principal and (2) the *functional* or technical responsibility of meeting given standards, such as capable and efficient conduct or the "right" course of action (Friedrich 1963, pp. 310-311). The first is a *dependent* kind of responsibility; the second is an *independent* responsibility. In the first sense, the representative is related to

somebody; in the second sense, reference is made to his "responsible conduct," thereby implying that the behavior of a representative is basically entrusted to his own conscience and competence.

When the distinction is translated into political terms, it appears that the expression "responsible government" covers two different expectations: (1) that a government is responsive and accountable to the people for what it does and (2) that a government behaves responsibly by acting efficiently and competently. We may call the first a *responsive* government and the second an *efficient* government. And the difference between the two is no small difference.

In our private dealings it matters little whether we lean more on personal or on functional responsibility, for in either case the representative has only one end in view: the sole interest of the principal, whatever happens to any other interest. But when political representation is concerned, another end is at stake: the interest of the whole, whatever happens to private interests. For this reason, the distinction between dependent and independent responsibility is, in politics, a crucial distinction, and it matters very much which kind of responsibility a representational system leans on. It is precisely on account of its independent margin of functional responsibility that a government is empowered to pursue the interest of the whole against partial and partisan interests. The more that independent responsibility gives way to dependent responsibility, the more likely it is that the general interest will give way to the claims of sectional, short-sighted, and contradictory interests.

It is not a paradox, then, to say that a responsible government can also be a very irresponsible government. Actually, the more responsive it becomes, the less it is in a position to behave responsibly. Ultimately, a choice must be made between the representativeness and the efficiency of government. We cannot ask a government to do one thing and its very opposite, that is, to yield to the demands of the governed and resist the "irresponsible demands" of the governed. More exactly, we cannot have, at the same time, *more* responsiveness and *more* independent responsibility. [See RESPONSIBILITY.]

### Types of representational systems

The foregoing explains why representational systems belong to two main patterns originating in England and France. The English type of representational system is based on a single-member plurality system of electoral representation, which affords little electoral choice and favors, therefore,

a two-party system; whereas the French type is based on proportional systems of electoral representation, which afford a large electoral choice and thereby facilitate multiparty systems. The English type sacrifices the representativeness of parliament to the need of efficient government, while the French type sacrifices efficient government to the representativeness of parliament.

If the vocabulary of politics were updated, the English type of representational system could fittingly be labeled "cabinet system" and the French type could properly be called "parliamentary government." Whatever the terminology, "representative government" poses a two-sided problem: governing and representing. The English (and American) type of political system maximizes the governing requirement, whereas the French (and Continental) type has maximized the mirroring requirement.

To be sure, in a number of countries one finds a more balanced combination of efficient government and representative representation (for example, the countries having both proportional representation and a relatively limited number of parties, say from three to five). Nonetheless, from the viewpoint of institutional engineering the fact remains that we cannot build a representational system that maximizes at one and the same time the function of functioning and the function of mirroring. There is a point at which we must choose, and the realistic alternatives are independent or dependent responsibility, rather than governed and governing democracy (Burdeau 1949-1957, vol. 6), or feigned and veritable self-government (Ross 1955, pp. 50-51).

The more developed and differentiated representational systems of the Western nations do not exhaust the range of all possible representational systems. Yet to the extent that the political structures of the developing nations are of a derivative nature, these nations are likely to be confronted with the issues and the problems that we have been discussing. The experience of the West does provide, therefore, a parameter for stating the minimal requirements for the existence of any representational system.

A representational system cannot materialize unless electoral accountability makes the governors responsible to the governed. However, all that is required is personal or dependent responsibility, since functional or independent responsibility qualifies the performance, not the relationship. Even the requirement of dependent responsibility should not be taken too literally; it merely implies *responsiveness*, albeit a sanctionable and safeguarded responsiveness.



In conclusion, a political system qualifies as a representational system whenever honest electoral practices secure a fair amount of responsiveness of the governors to the governed. This does not necessarily imply a large enfranchisement; but it does imply that no representational system can be based merely on "virtual representation." Nor does a political system qualify as a representational system if it has one ruler (monarch or dictator) who monopolizes the claim of representing the whole. Unless the representative function is entrusted to some collective body that is free enough to express a diversity of views and interests, we may be confronted (at least existentially) with representative rulership but not with a representational system.

### Contemporary issues

**Scale and scope of representation.** When representational systems were introduced in England and on the Continent, the electorates were small and so were the governments. With the passing of time the electorate has grown from a few hundred to tens of thousands of electors for one representative. Correlatively the small government with simple problems and few functions has become a large government with complex problems and innumerable functions. Both developments combine to place an ever-growing strain on the relational aspect of representation, that is, on the link between those represented and their alleged representatives. As Bruno Leoni put it, "... the more numerous the people ... one tries to 'represent,' ... and the more numerous the matters in which one tries to represent them, the less the word 'representation' has a meaning referable to the actual will of actual people other than that of the persons named as their 'representatives'" (1961, p. 18).

While political representation has always been a many-to-one relationship, the many have become so numerous that one may well wonder if, in a 50,000-to-1 ratio, it still makes sense to say that each person is represented. The reply is related to the scope of representation and of representative government. As long as representation is considered basically a protective device that limits (among other things) the scope of government, the reply is still, "Yes, it surely makes sense."

But the wider the coverage and the larger the area of intervention that calls the representative to decide upon matters that fall beyond the understanding of the represented, the more one may be alarmed by the potential of arbitrariness that is involved and wonder whether the scope of representation can be stretched without end. At any rate, one can hardly escape the feeling that the

initial link in the chain, the represented, has become an infinitesimal quantity.

**"Who" is represented?** The problem may be stated as follows: The larger the size of the electorate, the smaller the scale of representation; and the smaller the scale of representation, the more we lose sight of *who* is being represented. At least, this seems an inescapable conclusion if the *who* of representation refers to the "individual."

While this conclusion is not always acknowledged in this form, it helps to explain the dissatisfaction with the individualistic or atomistic basis of the theory and practice of representation. The opposition to "individualistic representation" need not be explained as merely a nostalgic revival of medieval corporate representation. However, the fact is that whenever the anti-individualistic mood has prevailed (as in Italy and Germany during the 1930s), it has been used to justify an autocracy rather than to suggest new means or techniques of nonindividualistic representation.

Without abandoning the starting point of the "individual voter," we may scrutinize his voting behavior with reference to the representational message, or intention, that he transmits. It has been suggested that the act of voting expresses (1) what the elector *has to say* (or thinks), or (2) what the elector is (existentially), or (3) what the elector *wills*. According to the first view, representation represents opinion; according to the second, it represents class or vocational appurtenance; according to the third, an individual may be represented even though he is inarticulate or silent.

The first is the traditional view, and is apparently destined to succumb under the joint blows of electoral numbers and big government. Whether we like it or not, we are thus likely to fall back on the other two interpretations. Both make the individual voter *less* of an individual. For if we vote by identifying ourselves with a social class, the fact that we vote per capita does not mean that we vote qua individuals. And the "will theory" of representation not only can be conceived collectivistically but can, moreover, be put to use for any purpose whatever, since a will that is silent or not articulated can be said to wish anything that its interpreter thinks it should.

These two approaches provide a way out precisely because they no longer attempt to answer the query concerning *who* is being represented. They actually provide an answer concerning *what* is being represented—which is not quite the same thing.

**"What" is represented?** Since all representational systems adopt a territorial criterion of elec-

toral apportionment, what is actually being represented are the localities, the geographic areas. But acknowledgment of this only leads to a reformulation of the question, What comes to be represented through a territorial channeling? The reply is largely a matter of speculation (or of needed research), except for one definite conclusion: territorial representation does not satisfy and even impedes occupational or functional representation.

The question of what is being represented can also be approached by discussing whether representation is focused more on ideal preferences or on material interests, more on values or on wants. And while territorial representation considers the citizen (not the *homo economicus*), thereby implying, among other things, that the voter should be encouraged to respond in terms of ideals rather than wants, the contention is that the adoption of vocational representation would definitely elicit a response in terms of material interests. In any case it should be pointed out that whether or not we believe in "individualism," and whether or not territorial representation works as it is supposed to work, we are made to vote qua individual citizens according to a territorial criterion of apportionment because it is the safest way of getting the electoral business done. If territorial apportionment is already exposed to tricky manipulations ("gerrymandering"), vocational or any other kind of functional apportionment would allow no end of manipulative trickery. [See APPORTIONMENT.]

In free societies, the idea of replacing a political parliament with an occupational parliament (or parliament of experts) has only led to attempts to combine the two. Side by side with the political parliament, some kind of technical or vocational parliament was established expressing the views of special interests, for example, the *Reichswirtschaftsrat* of the Weimar Republic and the economic advisory councils established in the 1950s in Italy and France (but not in West Germany). The attempt has definitely failed in Italy while it has been relatively successful in France and in some smaller countries. Ultimately, however, great difficulties arise out of the relative positions of the technical and the political parliaments. There are three possibilities: either (1) the technical parliament has the final say, or (2) the two parliaments are made equal, or (3) the technical parliament is a subordinate consultative body. The first case would in effect eliminate the political parliament; the second would lead to systematic conflict and paralysis; whereas in the third instance the consultative body would be listened to only when the political parliament was in the mood to do so.

What is being represented, then, in or by a political parliament? Local and community interests, class appurtenance, special and sectional interests, ideals, individual wants? Depending on the scale of representation, different combinations of all of these are represented. All sorts of voices that are strong enough to make themselves heard somehow find their way to a representative body.

The "how" of representation. Whatever the "who" or the "what" of representation, there remains the problem of the "how." And, clearly, the how of representation feeds back on the what and even on the who.

Broadly speaking, the how refers to the way in which a representational system is established and made to operate. Precisely who and what benefits from a given representational system may remain obscure, but one thing is clear: if the how is known, a system of elective representation (as defined above) allows more freedom and provides more responsiveness than any other known system.

More specifically, the how of representation can be taken to mean the style of representation, as Wahlke and his associates (1962) call it. In this sense the how of representation is related to the role orientations of the representative, for example, whether he acts more as a delegate or more as a trustee or whether he is more a party man, a constituency servant, or a mentor (Eulau 1962).

From another point of view, the how of representation is related to the electoral system and to the party system. The relationship between electoral systems and representational systems has already been mentioned and is a relatively old problem (for example, Duverger 1951). A more recent problem is how the representational process is affected by party interposition. For the more mass democracy there is and the more mass parties there are in a political system, the more the how of representation hinges on the party system as a vehicle of the representational process.

Party representation. The scale of representation being as it is, parties are a means for reducing unmanageable numbers to a manageable size. Citizens in modern democracies are represented *through* and *by* parties. This seems inevitable. The troubling thing is, however, that a point may be reached, or even has been reached, at which, as Beer (1965, p. 88) concisely put it, "the function of representing the national interest, once attributed to the Sovereign and later to Parliament, is now performed by party. Party—to echo Herman Finer's phrase—is indeed 'king.'"

Now, to speak of the party as a "filter" for political representation is one thing; to speak of the



party as "king" is quite another thing. The theoretical and constitutional problems that have resulted from this development are thorny; this is one of the reasons why even the more recent constitutions avoid giving formal recognition to party representation (the exceptions are the constitution of Brazil, the Bonn constitution, and the French constitution of 1958).

A realistic view of the present-day representational process confronts us, then, with a two-step process or even with a process that is cut in two: a relationship between the electors and their party, and a relationship between the party and its representatives. And the contention is that party nomination, that is, party co-optation, tends to be the real election; for while the electors choose the party, the elected are actually chosen by the party.

Of course parties, party systems, and countries are very different, and therefore generalizations can be very misleading. Yet it can be argued that whenever we are confronted with rigidly and powerfully organized mass parties—as is usually the case in Europe—the representative is likely to be far more a spokesman for his party than for anyone else (including his electors), and party bonds are perhaps far stronger than any other bonds (including the ties of social class). Thus, according to Duverger (1956), the modern representative is entrusted with a "double mandate," from his electors and from the party; and in practice the party mandate takes precedence over the electoral mandate.

Representation has lost all immediacy and can no longer be viewed as a *direct* relationship between the electors and the elected. The process hinges on three elements—those represented, the party, and the representatives. And the intermediary link appears to be so decisive that a point could be reached at which parliamentary representation would "resemble" party personnel, in the strict meaning of party career men, much more than the society it was asked to resemble. If so, the party would indeed be king and one could conclude that the *who* being represented is actually the party. In spite of the lack of adequate research, the available evidence suggests that this point has not yet been reached, although the parliamentary duplication of professional party politicians is progressing. [See PARTIES, POLITICAL.]

Although not unusual, it is perplexing that while actual political developments point in one direction, much of the prevailing theory and practice point in another. The scale of representation has become infinitesimal; the scope of representation

bypasses the grasp of the common man; furthermore, parties have largely replaced the electorate in deciding what should be represented and how. All these developments seem to indicate that what is really at stake is "responsibility" and that the real problem is how to improve the performance of big government in terms of independent responsibility without too great a loss of accountability. Yet much of the recent literature still emphasizes the sociological approach to representation, and proportional representation still enjoys the reputation of being the one truly representative electoral system. Since both concerns imply "similarity," or representativeness, they could be described as somewhat anachronistic.

### Research issues

To the extent that representational systems are normative constructions, they are largely beyond empirical disproof. Representational systems also presuppose, however, factual assumptions that can, and should, be tested empirically. In this connection there are three main areas of inquiry that are relevant to the theory of representation: (1) public opinion, that is, what people really want and expect; (2) decision makers, that is, who is actually chosen (how) to serve as a representative; and (3) representational behavior [see REPRESENTATION, *article on* REPRESENTATIONAL BEHAVIOR].

Despite field surveys and opinion polls, we do not know as much as we should concerning what people really demand from a representational system. According to one author, "the English people still like being governed," and the evidence suggests that the order of priorities is, for the British, the following: (1) consistency, prudence, and leadership; (2) accountability to Parliament and the electorate; (3) responsiveness to public opinion and demands (Birch 1964, p. 245). Possibly this could also be the order of priorities in West Germany, while perhaps in the United States responsiveness ranks second. On the other hand, one might guess that in countries like France (during the Third Republic and Fourth Republic) and Italy, the order of priorities would be (1) responsiveness, (2) accountability, and (3) leadership (that is, independent responsibility). So far, the results of only one nation-wide study are available and they concern constituency influence in the U.S. Congress (Miller & Stokes 1963).

As to the second focus of inquiry, decision makers in the legislative arena, there is growing documentation concerning *who is who* in parliaments (Meynaud 1967). While the data are very uneven and difficult to compare, we are now in a

position to test at least some aspects of the "resemblance" theory of representation on a country-by-country basis: England (Guttsman 1961; Berrington & Finer 1967), France (Dogan 1961; Hamon 1967), United States (Matthews 1954a; 1954b; 1967; Miller & Stokes 1965), Italy (Sartori 1957; 1967), Israel (Akzin 1967).

The sociological approach to representation that is implied by the foregoing documentation is seldom focused, however, on the problem of party representation: for example, the parliamentary duplication or projection of party career personnel has as yet received very little attention. In general, the role of the party as an intervening variable has remained a neglected area of research. The findings on whether (and if so to what extent) it is the party that administers the sanction of removal, or on whether the party is the real recruiter, are poor, peripheral, and widely conflicting (for example, Seligman 1958, p. 361; Sorauf 1963, p. 120). Likewise, the claim that parties receive an "electoral mandate" remains a matter of speculation, even though it could be warranted or disclaimed by ascertaining whether this is really the expectation of the voter.

The resemblance aspect of representation has been explored far more than the aspects of accountability and responsibility. Possibly research in that area is easier. This is all the more reason for stressing that a reorientation of the focus of research on accountability is badly needed. For instance, whose sanction of removal is feared most—the sanction coming from the electorate, from the party machine, or from other supporting groups? This is a rather crucial question—and a question that remains unanswered.

GIOVANNI SARTORI

[See also ELECTIONS; LEGISLATION; PARTIES, POLITICAL; POLITICAL EXECUTIVE. Other relevant material may be found in GOVERNMENT; POLITICAL PARTICIPATION.]

#### BIBLIOGRAPHY

A penetrating analysis of the issues raised by political representation in the context of Continental constitutionalism, with particular reference to France, may be found in Carré de Malberg 1922. Sartori (1957) 1962 includes a survey of the literature; Birch 1964 covers the British approach; and de Grazia 1951 is relevant for the American pattern. Concerning the medieval origins of representation, Clarke 1936 is still a valuable analysis, in addition to the classic Carlyle & Carlyle 1903–1936. With reference to the European constitutional development of the theory of representation, the most influential work has been Gierke 1868–1913.

AKZIN, BENJAMIN 1967 The Knesset. In Jean Meynaud (editor), *Decisions and Decision Makers in the Modern State*. Paris: UNESCO.

- BAGEHOT, WALTER (1859) 1965 *Parliamentary Reform*. Pages 296–347 in Walter Bagehot, *Bagehot's Historical Essays*. Garden City, N.Y.: Anchor.
- BEER, SAMUEL H. 1965 *British Politics in the Collectivist Age*. New York: Knopf.
- BERRINGTON, H. B.; and FINER, S. E. 1967 The British House of Commons. In Jean Meynaud (editor), *Decisions and Decision Makers in the Modern State*. Paris: UNESCO.
- BIRCH, ANTHONY H. 1964 *Representative and Responsible Government*. London: Allen & Unwin.
- BURDEAU, GEORGES 1949–1957 *Traité de science politique*. 7 vols. Paris: Librairie Général de Droit et de Jurisprudence. → See especially Volume 6.
- CARLYLE, ROBERT W.; and CARLYLE, A. J. (1903–1936) 1950 *A History of Mediaeval Political Theory in the West*. 6 vols. New York: Barnes & Noble; London: Blackwood. → See especially Volume 5, Chapter 9, and Volume 6, Part 2, Chapter 6.
- CARRÉ DE MALBERG, RAYMOND (1922) 1962 *Contribution à la théorie générale de l'état: Spécialement d'après les données fournies par le droit constitutionnel français*. Volume 2. Paris: Sirey.
- CLARKE, MAUDE V. (1936) 1964 *Medieval Representation and Consent: A Study of Early Parliaments in England and Ireland, With Special Reference to the Modus tenendi parliamentum*. New York: Russell.
- DE GRAZIA, ALFRED 1951 *Public and Republic: Political Representation in America*. New York: Knopf.
- DOGAN, MATTEI 1961 Political Assent in a Class Society: French Deputies 1870–1958. Pages 57–90 in Dwaine Marvick (editor), *Political Decision-makers*. New York: Free Press.
- DUVERGER, MAURICE 1951 The Influence of the Electoral System on Political Life. *International Social Science Bulletin* 3:314–352.
- DUVERGER, MAURICE 1956 Esquisse d'une théorie de la représentation politique. Pages 211–220 in *L'évolution du droit public*. Paris: Sirey.
- EULAU, HEINZ 1962 The Legislator as Representative: Representational Roles. Pages 267–286 in John Wahlke et al., *The Legislative System: Explorations in Legislative Behavior*. New York: Wiley.
- FRIEDRICH, CARL J. 1962 Representation. Volume 19, pages 163–167 in *Encyclopaedia Britannica*. Chicago: Benton.
- FRIEDRICH, CARL J. 1963 *Man and His Government: An Empirical Theory of Politics*. New York: McGraw-Hill.
- GIERKE, OTTO VON (1868–1913) 1954 *Das deutsche Genossenschaftsrecht*. 4 vols. Graz (Austria): Akademische Druck- und Verlagsanstalt. → Volume 1: *Rechtsgeschichte der deutschen Genossenschaft*. Volume 2: *Geschichte des deutschen Körperschaftsbegriffs*. Volume 3: *Die Staats- und Korporationslehre des Altertums und des Mittelalters und ihre Aufnahme in Deutschland*. Volume 4: *Die Staats- und Korporationslehre der Neuzeit*.
- GIERKE, OTTO VON (1881) 1958 *Political Theories of the Middle Age*. Cambridge Univ. Press. → First published as "Die publicistischen Lehren des Mittelalters," a section of Volume 3 of Gierke's *Das deutsche Genossenschaftsrecht*. Translated with a famous introduction by Frederic William Maitland.
- GIERKE, OTTO VON (1913) 1934 *Natural Law and the Theory of Society: 1500 to 1800*. Translated with an introduction by Ernest Barker. 2 vols. Cambridge Univ. Press. → A translation of five subsections of



- Volume 4 of *Das deutsche Genossenschaftsrecht*. A paperback edition was published in 1957 by Beacon.
- GOSNELL, HAROLD F. 1948 *Democracy: The Threshold of Freedom*. New York: Ronald Press.
- GUTTSMAN, W. L. 1961 Changes in British Labour Leadership. Pages 91-137 in Dwaine Marvick (editor), *Political Decision-makers*. New York: Free Press.
- HAMON, LÉO 1967 Members of the French Parliament. In Jean Meynaud (editor), *Decisions and Decision Makers in the Modern State*. Paris: UNESCO.
- Il parlamento italiano, 1946-1963*. A research project directed by Giovanni Sartori. 1963 Naples: Edizioni Scientifiche Italiane.
- KELSEN, HANS (1945) 1961 *General Theory of Law and State*. New York: Russell.
- LABAND, PAUL (1876-1882) 1911-1914 *Das Staatsrecht des deutschen Reiches*. 5th ed. 4 vols. Tübingen (Germany): Mohr.
- LA BIGNE DE VILLENEUVE, MARCEL DE 1929-1931 *Traité générale de l'état: Essai d'une théorie réaliste de droit politique*. 2 vols. Paris: Sirey.
- LEIBHOLZ, GERHARD (1929) 1960 *Das Wesen der Repräsentation und der Gestaltswandel der Demokratie im 20. Jahrhundert*. 2d rev. ed. Berlin: Gruyter.
- LEONI, BRUNO 1961 *Freedom and the Law*. Princeton, N.J.: Van Nostrand.
- MATTHEWS, DONALD R. 1954a *The Social Background of Political Decision-makers*. Garden City, N.Y.: Doubleday.
- MATTHEWS, DONALD R. 1954b United States Senators and the Class Structure. *Public Opinion Quarterly* 18:5-22.
- MATTHEWS, DONALD R. 1967 United States Senators: A Collective Portrait. In Jean Meynaud (editor), *Decisions and Decision Makers in the Modern State*. Paris: UNESCO.
- MEYNAUD, JEAN 1967 Introduction. In Jean Meynaud (editor), *Decisions and Decision Makers in the Modern State*. Paris: UNESCO.
- MEYNAUD, JEAN (editor) 1967 *Decisions and Decision Makers in the Modern State*. Paris: UNESCO. → Earlier versions of the articles in this book were published in Volume 13 of the *International Social Science Journal*.
- MILLER, WARREN E.; and STOKES, DONALD E. 1963 Constituency Influence in Congress. *American Political Science Review* 57:45-56.
- MILLER, WARREN E.; and STOKES, DONALD E. 1965 *Representation in Congress*. Englewood Cliffs, N.J.: Prentice-Hall.
- ORLANDO, VITTORIO E. (1895) 1940 Del fondamento giuridico della rappresentanza politica. Pages 417-456 in Vittorio E. Orlando, *Diritto pubblico generale: Scritti vari (1881-1940) coordinati in sistema*. Milan: Giuffrè.
- ROSS, JAMES F. S. 1955 *Elections and Electors: Studies in Democratic Representation*. London: Eyre & Spottiswoode.
- SARIPOLOS, NIKOLAUS N. 1899 *La démocratie et l'élection proportionnelle*. Paris: Rousseau.
- SARTORI, GIOVANNI (1957) 1962 *A teoria da representação no estado representativo moderno*. Belo Horizonte (Brazil): Edições da Revista Brasileira de Estudos Políticos.
- SARTORI, GIOVANNI 1967 Parliamentarians in Italy. In Jean Meynaud (editor), *Decisions and Decision Makers in the Modern State*. Paris: UNESCO.

- SELIGMAN, LESTER G. 1958 Party Roles and Political Recruitment. *Western Political Quarterly* 11:359-361.
- SORAU, FRANK J. 1963 *Party and Representation: Legislative Politics in Pennsylvania*. New York: Atherton.
- WAHLKE, JOHN et al. 1962 *The Legislative System: Explorations in Legislative Behavior*. New York: Wiley.

## III

## REPRESENTATIONAL BEHAVIOR

Representational behavior refers to that aspect of legislative conduct that is studied for the nature, extent, and basis of its correspondence with constituency opinions. Representatives are assumed to differ in the extent to which they reflect their constituents' opinions in their legislative activities, most notably in their decision making. The modern student of representational behavior is interested in (1) determining the extent of agreement between constituency opinion and legislative behavior on various issues, (2) accounting for differences observed in the relationships between representatives and their constituencies, and (3) specifying the effect of different representative-constituency relationships upon public policy formation.

The study of the representational behavior of individual legislators is particularly important in political systems where party voting is not essential for governmental stability. In a parliamentary system featuring well-disciplined parties and slim majorities, most notably the modern British Parliament, individual representatives are subject to restraints on their voting behavior that are unknown in other legislative bodies, such as the United States Congress, where members are relatively free to vote their opinions or to respond to political forces other than the party leadership. Although eighteenth-century Britain can be credited with the birth of interest in representative-constituency relationships, the United States has produced the most extensive research on the topic. In Britain, within the last century, the motivations for research on representational behavior apparently decreased as party voting increased in parliament.

**Development of interest.** The study of representational behavior and representative-constituency relationships assumed importance with the development of representative democratic government. Democratic theory held that the general public ought to have some participation in governmental policy making so that the government would do what the people wanted it to do (or at least did not do what they did not want). All modern democracies eventually developed some form of representative government, which institutionalized popular participation in governmental policy making

through popular election of representatives empowered to formulate governmental policy. The establishment of a popularly elected legislative body did not, however, insure that governmental policy would follow public opinion. Once elected, the representatives might or might not satisfy popular desires in formulating public policy.

Viewed strictly from the standpoint of democratic theory, it was undesirable for representatives to act contrary to public opinion. But at the same time, it was obvious that complete reliance upon public opinion in governmental policy making was also likely to have undesirable consequences. The voters' ignorance or misunderstanding of the facts involved in the issues, their inclinations toward selfishness in promoting their own interests, and their tendencies toward hasty and ill-considered actions were cited as factors that would lead to unwise or unjust governmental policies if the representatives were merely to act in accordance with their constituents' wishes.

On one hand, then, both abstract requirements of democratic theory and practical consequences of making representatives directly accountable to the people urged a close correspondence between constituency opinion and legislative behavior. On the other hand, both abstract notions of desirable government and practical necessities of rational decision making argued *against* a close correspondence between them. The conflict between these considerations was of concern to scholars who sought to evaluate the operation of representative government. This conflict provided the initial interest in the study of representational behavior.

Those who were interested primarily in evaluating the operation of representative government sought to determine the proper relationship between the representative and his constituency. Later students who were less interested in the evaluative aspects of political science were also less concerned with determining the "proper" relationship between the representative and his constituency. They sought instead to clarify types of relationships, discover the factors associated with different relationships, and predict the probable consequences of different relationships for public policy.

**Early approaches.** Edmund Burke's defense of his voting record in his classic "Speech to the Electors of Bristol" in 1774 set forth a conception of the proper role of the representative as one who ought to respect his constituents' opinions, who ought to prefer their interests above his own, but who ought *not* to sacrifice his unbiased opinions in deciding for the good of the whole nation

(quoted in de Grazia 1951, p. 38) [see BURKE]. Burke's often quoted statement most certainly influenced the subsequent study of representative-constituency relationships. Inquiry into the topic commonly raised the general question to which Burke addressed himself: What ought to be the relationship of the representative to his constituency? This question elicited normative prescriptions but did not invite empirical inquiry, and the literature that developed was characterized by opinions and arguments concerning the proper role of the representative. In general, extreme democrats (e.g., the Levellers of seventeenth-century England and the "direct democrats" in the early twentieth-century United States) insisted that the representative ought to obey the wishes of his constituents, while those less enamored with the unqualified application of democratic principles argued that better legislation would result if the representative relied mainly on his own judgment when deciding upon legislation.

Interwoven with the question of whether the representative ought to obey the wishes of his constituents or vote according to his own judgment was another question: Ought the representative to be guided in his decisions by the welfare of his constituency or the welfare of the state as a whole? The logical independence of these questions was not always recognized in the polemics concerning the proper role of the representative. The distinction between the two questions has recently been acknowledged and labeled as a difference between the "style" and the "focus" of representation (Wahlke et al. 1962, p. 269).

The style of representation refers to the particular criterion of judgment the representative ought to use in deciding on legislative issues. The basic argument over the proper style of representation was whether the representative ought to behave as a "trustee" and arrive at decisions either on the basis of his own sense of right and wrong or on his personal evaluation of the facts or whether he ought to behave as a "delegate" and disregard his personal opinions while executing those of his constituents. (It should be noted that the trustee-delegate terminology is not standard in the literature. Some writers have reversed the application of the terms or substituted other labels, such as "agent" or "ambassador," for those given here.)

The focus of representation refers to the particular group of persons whose welfare the representative ought to consider in deciding on legislative issues. The basic argument over the focus of representation was whether the representative



ought to be "district-oriented" and consider primarily the welfare of his constituents or whether he ought to be "nation-oriented" and consider the welfare of the nation as a whole.

Argument over the normative question of how the representative ought to act characterized the literature on representative-constituency relationships from the second half of the eighteenth century to the middle of the twentieth century. Like many other "ought" questions in political science, the question of the proper role of the representative defied definitive answers. The answer to this question involved more than reasoned appeals to value positions concerning what constituted good government. Vital information was lacking on empirical aspects of representative-constituency relationships, such as the nature and extent of the constituents' knowledge of legislative issues, the ways by which the constituents' opinions on issues were communicated to the representative, the correspondence between his perceptions of constituents' opinions and their actual opinions, the politics of his district, and so forth. It became obvious that detailed knowledge of these and other factors was not only important in its own right but was also essential for those who remained interested in the classic normative question.

**Recent developments.** Modern research in representational behavior has been characterized by its concern for empirical questions, but research to date has been handicapped by the unavailability of certain behavioral and psychological data required for an adequate understanding of the subject. In order to determine the faithfulness with which representatives represent their constituents' wishes, for example, data are needed both on the behavior of incumbents and on the opinions of their constituents. In general, relevant data on the representatives' behavior has been more accessible than matching data on their constituents' opinions.

Research on representational behavior has been aided by the ready availability of public information about important aspects of the incumbents' behavior, such as votes cast on the floor and sponsorship of bills. Information on other legislative activities, such as speaking on proposed bills, may also be found in published sources or observed directly. Behavior in committee meetings, party conferences, or outside the legislature proper is less easily learned but may be determined through interviews with representatives or other participants in the political process. Despite the wide variety of legislative activities that could have relevance for the disposition of public issues on which constituents have opinions, studies have generally relied

solely on the representative's floor votes for the data on his side of the relationship. Although some doubts have been raised concerning the wisdom of relying on the recorded floor vote as the sole indicator of the representative's behavior, there seems to be little question that floor votes do constitute important data. In legislative bodies that feature strong party voting, however, some aspects of a representative's activities other than his floor votes would probably be more suitable for research.

The overriding research problem has not been obtaining data on representatives' behavior but on constituents' opinions. The sample survey is the most suitable research technique yet developed for obtaining reliable information about the distribution of personal opinions among populations. An appropriately drawn sample of about 2,000 respondents furnishes an adequate basis for making confident statements about the distribution of voters' opinions in the United States. But the science of probability sampling is such that a sample of practically the same size is needed to make statements at a similar level of confidence about a population of only 400,000—about the size of the average congressman's constituency in the United States. Because careful sample surveys are expensive to conduct and complex to administer, survey techniques until very recently have not been used at all for systematic research in representational behavior.

A pioneering effort was undertaken in 1958 by the University of Michigan's Survey Research Center, which designed its national sample of the United States electorate in a manner that permitted limited analysis of constituency opinions in about one hundred congressional districts (Stokes & Miller 1962). That effort opened the door for a genuine breakthrough in the study of representational behavior, enabling the first direct comparisons to be made between representatives' behavior and their constituents' opinions (Miller & Stokes 1963). Much additional work remains to be done with this approach, however, before firm generalizations can be made concerning the interesting empirical questions of representative-constituency relationships. It is unlikely that these generalizations will soon be formulated, for cost and complexity factors involved in sampling a number of constituencies still obstruct the development and application of this methodological approach.

Without the use of sample survey methods, students of representational behavior have conducted their research by classifying constituencies according to social characteristics on which census-type data were available (such as occupation, per cent

urban, per cent native-born, and so on) and then correlating these characteristics with the voting positions of their representatives. In some cases, researchers have attributed opinions to the inhabitants of certain districts on the basis of their social composition: constituents in working-class districts were said to favor unemployment guarantees, those in farming districts were said to favor agricultural subsidies, and so forth. In other cases, this somewhat risky imputation of opinions has been avoided, and researchers have simply examined similarities and differences in the voting patterns of representatives from various types of districts. A common refinement of such research involved introducing the variable of party competition in the district and analyzing legislators' votes according to the "safe" or "doubtful" nature of their constituencies as determined by past election returns.

Studies of roll call voting in the United States Congress and in American state legislatures have disclosed tendencies for representatives from different types of districts to vote on certain measures in accordance with presumed or imputed interests of their constituencies. The strength of these tendencies appears to increase with an increase in homogeneity of the social composition of the district. Those representing districts composed almost entirely of working-class people, for example, tend to give more support to legislation establishing unemployment guarantees than those representing districts dominated by working-class people but with substantial white-collar elements. Nevertheless, constituency characteristics have been shown to account for only a part of voting behavior in American legislatures. The representatives' party affiliations generally stand out as the most pervasive influence on voting alignments, despite the undisciplined nature of American parties. The importance of party affiliation is so clearly recognized that researchers often attempt to measure constituency influences on voting by examining the instances when representatives abandon the party position. When the party's position on an issue coincides with the position assumed to be favored by the constituency, the representative's vote is unlikely to be contradictory. When the party's position and the assumed constituency position are in conflict, his vote is less predictable. There is some evidence to indicate that a representative will tend to favor the presumed interest of the constituency in case of such conflict if he comes from a "doubtful" rather than a "safe" district.

Students of representational behavior who have made use of census-type data on constituency characteristics in lieu of sample survey data on consti-

tuents' opinions have produced these and other suggestive findings. But they have not been able to answer some of the most important empirical questions because they have lacked reliable knowledge about the nature and distribution of constituency opinions.

In the absence of survey information about constituents' preferences, the exact impact of constituency opinions on representational behavior cannot be measured. Not only is the researcher unable to establish the extent to which the legislator's behavior conforms to constituency opinions, but he is also unable to determine how closely the representative's personal attitudes and perceptions match these opinions.

**Research needs.** The two basic types of data required for research in representational behavior are, for reasons mentioned above, data on constituency opinions and data on representatives' behavior. At least four other categories of information, which remain to be discussed, are especially important for the study of representational behavior. Data are needed on representatives' personal attitudes toward policy issues, the motivational bases of their behavior, their perceptions of their constituencies, and representative-constituency communications. Each of the last four categories will be discussed in turn.

*Representatives' personal attitudes.* In general, students of legislative behavior have shown that a representative's voting record reflects, to some extent, his personal attitudes as determined by independent attitude tests. Representatives whose voting behavior suggests that they are especially attuned to the opinions of their constituents may in fact exert little or no effort to ascertain district opinions and may merely vote according to their own personal attitudes. Their votes, however, may conform closely to their constituencies' positions simply because they hold the same basic attitudes and values as their constituents. The mechanics of representation operating in this situation are quite different from those involved when a close correspondence between constituency opinions and voting behavior exists regardless of the representative's personal attitudes toward the issues. Shifts in constituency opinions over time are more likely to be reflected in the representative's behavior when he is consciously responsive to those opinions. In order to assess the operation of constituency control over a representative's behavior, it is necessary to know something of his personal attitudes toward the issues on which he votes.

*Motivational bases.* In order to understand why a representative acts in a given way, it is necessary



to give explicit attention to the motivational forces guiding his behavior. Scattered attention has been given to motivational factors in case studies of legislative decisions, but this obviously important variable has been largely ignored in more comprehensive studies of legislative behavior. Undertaking motivational explanations of human behavior may appear to be hopelessly complex. But complexity can be avoided and the understanding of representational behavior can be helped considerably if some simple motivational concepts are introduced into the analysis. Three different motivational bases seem to be particularly relevant to the study of representational behavior. The representative's behavior might be guided by a desire to (1) express his own personal attitudes toward the legislation before the chamber, (2) fulfill his personal conception of what the proper representative *ought* to do (e.g., act as a delegate or a trustee), or (3) be re-elected to office at the next election. Probably no representative's behavior is completely guided by any one of these three motivational bases, and more than one base may underlie any given act. There undoubtedly are other important motivational bases. The point is, however, that the study of representational behavior will remain incomplete unless explicit attention is given to the collection and analysis of data relevant to the motivational bases of political conduct.

*Representatives' perceptions of the constituency.* A person's reactions to a situation depend on how he perceives it, and his perceptions do not necessarily accord with reality. It is important to determine the accuracy of the representative's perceptions of his constituents' opinions on various policy matters. The representative who consciously tries to vote in accordance with his constituents' opinions, but who frequently misreads them, may seem to be ignoring his district when in fact his behavior is due to a failure in communication instead of a lack of concern for constituency opinions.

It is also important to gather information on the representative's perceptions of the important political forces in his district, which can affect his chances for renomination or re-election. These forces may exert themselves at either stage, depending on the structure of politics in the district. They may, for example, be important at the nomination stage, where the official party endorsement is acquired. In some cases, the representative himself may control his own renomination through his domination of the district party organization. (This situation is not uncommon in the United States but perhaps occurs less frequently in countries with

stronger national parties.) In other cases, effective control at the nomination stage may lie with a small group of party leaders, whom the incumbent must court in order to be renominated. Finally, where the nomination is won through elections at a party primary or in conventions, there is the possibility that a large number of party workers or rank-and-file voters may be the force to be courted.

The importance of the general election stage as an obstacle to re-election depends on the tendencies toward party voting in the district. In districts that are safe for one party, failure to be renominated constitutes the only real threat to re-election. In competitive districts, the key to victory may involve cultivating support among various groups of voters. A candidate is not likely to be unaware of these obvious political facts, but because of his closeness to the situation or because of his anxiety over re-election, he may perceive them somewhat differently than an outside observer.

Data on the representative's perceptions of the district's politics are important for understanding his reactions to constituency-based demands, especially if he is strongly motivated by a desire for re-election. Data on the legislator's perceptions of his constituents' opinions have broad utility for the study of representational behavior but are especially important if the representative is strongly motivated to do what his constituents want, for then his behavior depends on how he perceives those wants.

*Representative-constituency communications.* The representative's perceptions of his constituency are affected by the content of the messages he receives and by the manner of their transmission. A useful distinction can be made between representative-initiated and constituency-initiated communications. The representative may himself try to determine the opinions of his constituents by examining the local newspapers, conducting his own public opinion polls, or circulating among the voters in his district. The amount of effort he expends in initiating communications with his constituents, the types of communication activities employed, and the specific constituents or constituency groups contacted can be expected to depend on the motivational basis of the representative's behavior and on his various constituency perceptions.

Constituents may also initiate communications to their representatives by such means as writing letters, sending telegrams, and making personal visits. Studies have shown that the opinions conveyed through various forms of constituency-initiated communications are unlikely to be accurate reflections of the distribution of constituency opinions due to the operation of distorting influ-

ences. For example, representatives are more likely to hear from those who agree with them than from those who disagree, and letters are more likely to be written in opposition to proposals than in support of them. Those who do initiate communications also tend to rank higher in education, income, political involvement, and occupational status than those who do not.

Sources of distortion are also present in representative-initiated communications. Legislators tend to communicate more with those who agree with their views, who are more highly involved in politics, and who share the same party identification. Although faithful disclosure of constituency opinions can be produced by professional polling or survey research organizations, few representatives can afford such expensive means of acquiring this information. Lacking essential technical and organizational resources of professional research firms, those who engage in opinion polling seldom develop accurate knowledge of constituency opinions because of sampling difficulties or biased wording of questions. The fact is that some representatives are relatively unconcerned about the accuracy of their polling procedures, for they frequently conduct polls for purposes of public relations with little intention of studying the findings for decision-making purposes.

Whether representatives poll their constituents for public relations or for decision-making purposes probably depends, once again, on their perceptions of the constituency and on their motivational basis of behavior. A delegate-styled representative, who is motivated primarily by a desire to perform his representational role by voting the way his constituents want, would probably strive for greater accuracy in polling and give more consideration to poll findings than would, for example, a trustee or a representative from a safe district whose chances for renomination are controlled by a group of party leaders. But regardless of the motivational bases of the representative's behavior, detailed information concerning the persons, groups, messages, and media involved in communications constitutes another important category of data for the study of representational behavior.

KENNETH JANDA

[See also INTEREST GROUPS; LEGISLATION; LOBBYING. Other relevant material may be found in DECISION MAKING; POLITICAL BEHAVIOR; POLITICAL RECRUITMENT AND CAREERS.]

#### BIBLIOGRAPHY

DE GRAZIA, ALFRED 1951 *Public and Republic: Political Representation in America*. New York: Knopf. → A

thorough summary and analysis of various views of the proper role of the representative.

DEXTER, LEWIS ANTHONY 1957 *The Representative and His District. Human Organization* 16:2-13.

MACRAE, DUNCAN 1958 *Dimensions of Congressional Voting: A Statistical Study of the House of Representatives in the Eighty-first Congress*. University of California Publications in Sociology and Social Institutions, Vol. 1, No. 3. Berkeley: Univ. of California Press.

MILLER, WARREN E.; and STOKES, DONALD E. 1963 *Constituency Influence in Congress. American Political Science Review* 57:45-56.

STOKES, DONALD E.; and MILLER, WARREN E. 1962 *Party Government and the Saliency of Congress. Public Opinion Quarterly* 26:531-546. → This is the first published report of the findings of a 1958 study conducted by the Survey Research Center, University of Michigan.

TURNER, JULIUS 1952 *Party and Constituency: Pressures on Congress*. Johns Hopkins Studies in Historical and Political Science, Series 69, No. 1. Baltimore: Johns Hopkins Press.

WAHLKE, JOHN et al. 1962 *The Legislative System: Explorations in Legislative Behavior*. New York: Wiley.

#### REPRESSION

See DEFENSE MECHANISMS; PSYCHOANALYSIS.

#### REPRISALS

See SANCTIONS, INTERNATIONAL.

#### REPRODUCTION, HUMAN

See FERTILITY; FERTILITY CONTROL; POPULATION.

#### REPUBLICANISM

See DEMOCRACY; MONARCHY.

### RESALE PRICE MAINTENANCE

In the practice of resale price maintenance the manufacturer determines and enforces the price at which distributors resell his product. Hence, resale price maintenance is also known as vertical price fixing, as price protection, or as the practice of imposed prices. The term "fair trade" is an American euphemism. A manufacturer may set resale prices for either or both the wholesale and retail stages of distribution, and the prices may be fixed prices or minimum prices.

The practice is confined largely to branded goods but is also found in unbranded products such as various building materials (for example, in the United Kingdom). Among branded goods, resale price maintenance is or has been associated with drugs, proprietary medicines and toilet preparations, books and magazines, phonograph records, domestic electrical appliances, furniture, motor vehicles, tobacco products, liquor, confectionery,



and office equipment. (The difficulty of controlling trade-in allowances has made price maintenance less effective in consumer durables.) The extent of the practice and the products involved differ greatly from country to country, being much affected by legal provisions. In the United Kingdom, where resale price maintenance has been extensive, it affected, at its peak, one-fifth or more of total consumer expenditure; in the United States the present proportion probably is well below 5 per cent.

**Effects on distributive trades.** Resale price maintenance, when it covers the bulk of a particular branch of trade, has the following effects on retailing (or, *mutatis mutandis*, on wholesaling): It eliminates or attenuates retail price competition, and it tends to make gross margins (the difference between selling prices and buying prices) and retail prices higher than they would be otherwise. Studies of the history of resale price maintenance and of its removal support this conclusion, as do statistical studies of prices in comparable states of the United States with and without price maintenance.

The removal of price competition directs retail competition into nonprice channels: the struggle for customers takes the form of the competitive extension of retail services, reflected in more expenditure on inventories, sales staff, provision of credit, amenities, and location. Nonprice competition tends toward a greater geographical dispersion of retail outlets and sustains a larger labor force in distribution. Competition of this kind ensures that the higher gross margins do not mean higher net returns on capital in retailing.

Resale price maintenance also tends to retard innovation and change in retailing. Historically, new forms of retailing have become established on the basis of competitiveness in price, achieved generally by a curtailment of retail services. Where resale price maintenance is effective and comprehensive, this way of entry is barred. The practice tends further to stabilize retailing in that it reduces the rate of growth of lower-cost firms by denying them the use of competitive price policies. It protects the less efficient, higher-cost firms.

The foregoing analysis helps to explain attitudes toward resale price maintenance. Historically, it generally owed its origins to the pressure of organized bodies of retailers (and wholesalers) anxious to avoid price competition, especially when threatened by new forms of retailing such as department stores and chain stores. Currently it enjoys its most vocal advocacy from such bodies of distributors. It also receives support from labor

unions of distributive workers. In some Western countries these supporters of resale price maintenance constitute a strong pressure group. On the other hand, it is rare for consumer groups to advocate resale price maintenance, although many of the claims made for the practice purport to show that it is in the consumer's interests.

**Effects on demand.** It may seem as if manufacturers of branded goods should oppose retail price maintenance, since its general effect in raising retail prices is against their interests. Nevertheless, many manufacturers seem to support the practice and use it where it is permitted and can be enforced. It must be that in some situations the adverse effect on sales of the higher retail price is (or is thought to be) more than offset by favorable consequences stemming (or thought to be stemming) from the practice. Several types of situations may be identified.

The most common situation—and historically the most important—is one in which the retailers favoring the control of retail price competition are able to influence materially the sales of particular brands (i.e., where together they can shift significantly the demand schedule of particular brands). In these conditions resale price maintenance secures the support of such retailers; the concomitant increase in their gross margins can be viewed as a toll the manufacturer has to pay for access to the established and effective network of retail outlets. But if the manufacturers of competing brands of a product all use price maintenance, the sales-promoting effects of the practice are neutralized; if each pays the toll, none has the advantage. Except in the special circumstances discussed below, the demand schedule for the product is unaffected; and therefore the higher retail prices are unfavorable to the manufacturers collectively.

The power of retailers favoring price maintenance depends on their relative strength (in terms of market share), their organization, their ability to influence consumer demand (through influencing the availability of the goods, their display, and point-of-sale advice or suggestions), and the absence of concerted action on the part of manufacturers to resist the retailers' pressure. In several countries the necessary combination of conditions was lacking in the grocery and clothing trades but present in the drug and liquor trades at the time when the price maintenance movement began. Again, resale price maintenance is relatively unimportant where innovating and price-competitive retailers have little difficulty in launching new brands or unbranded varieties of the goods they handle. Groceries and clothing may in this respect

be compared with books, automobiles, and cigarettes. The current long-term tendency toward impersonal methods of retail selling is likely to undermine resale price maintenance.

The influence of organized retailers is not the only reason why some manufacturers may find it profitable to have resale price maintenance. There are other ways, not dependent on dominant retailer support for price maintenance, in which the practice in certain circumstances may bring about favorable shifts of the demand schedule of a particular class of products, sufficient to outweigh the disadvantage of the increase in retail prices. While there is little difficulty in stating the conditions for an over-all favorable result for the manufacturers, it is not easy to discover whether these conditions are satisfied in any actual situation.

It is likely that consumer demand for some products is partly dependent on consumer access to the retail outlets that stock them and on information from retailers about them. Resale price maintenance stimulates sales insofar as assured bigger retail gross margins make possible and encourage more extensive and widespread stocking of the products and promote better-informed retail selling. It may seem, on the other hand, that in the appropriate circumstances—i.e., where the availability of the products and the provision of information are important in increasing demand—retailers have an incentive, even in the absence of resale price maintenance, to make the necessary investment in inventory and to incur expenditure on a knowledgeable sales force. But the incentive is weakened to the extent that consumers buy at lower prices from lower-cost shops after taking advantage of the facilities of higher-cost shops: retailers who provide the demand-creating services do not capture the full benefit of sales resulting from their outlays, which therefore tend to be reduced. (Incentive is also weakened to the extent—probably negligible in practice—that potential purchasers on impulse are deterred from buying by their knowledge that the price of the item in question is, or may be, lower elsewhere.) Resale price maintenance, by reducing the incentive for consumers to shop around for advantages, encourages the provision by retailers of the relevant before-sales services in the competitive search for customers. In such situations, the prevalence of which is debatable, manufacturers may wish to practice price maintenance. Resale price maintenance is, however, an indiscriminating method for encouraging retailers to follow desired inventory, display, and sales policies. In practice, it protects all retailers who are supplied with the prod-

ucts, irrespective of their trading policies, and stimulates the provision of services generally, not only presales services. Distribution through selected retailers is a more direct method for accomplishing the desired ends, and it can be achieved independently of resale price maintenance.

**Effects on manufacturer competition.** The conceivable advantages to be derived by the price-maintaining manufacturers in certain circumstances have so far been considered in terms of favorable shifts in the consumer demand schedule. Other possible advantages may arise where the consumer demand schedule is unaffected by the practice. Thus it is often implied that price competition in the retailing of their products would provoke more intensive competition among manufacturers and affect their selling prices to the trade and that, though their aggregate sales volume would be greater, profits would be smaller. This sort of effect is feared by many price-maintaining manufacturers, although it is difficult to determine whether these fears are justified in particular instances.

The transmission of price competition from retailing to manufacturing could follow one of two lines. First, retail price competition might lead to the concentration of more buying power in the hands of successful large retail enterprisers, who could use it to undermine stable trade-price structures. Second, price competition among retailers might cause variations in interbrand (retail) price relationships and hence in the relative sales of the various brands. Such changes in market shares might cause the disadvantaged manufacturing firms to reduce their trade prices in an attempt to regain lost business. These adjustments would inject instability into the structure of trade prices and might result in more intensive and persistent price competition. The second possibility may be restated in this way: resale price maintenance may be necessary or useful to underpin a manufacturers' horizontal price agreement or a noncollusive quasi agreement in tight oligopoly situations [see OLIGOPOLY]. The need for resale price maintenance is more apparent and more direct in those cases where manufacturers agree not to compete in price and reach the final consumers both directly and through independent distributors. The objectives of the agreement might not be fully achieved without the participation of the independent distributors, which could be secured or enforced by the imposition of resale prices. In this context, as in others, resale price maintenance is a partial substitute for vertical integration by manufacturers into distribution.



**Loss-leader selling.** No discussion of the economics of resale price maintenance is complete without some consideration of "loss-leader" selling. This ill-defined practice refers to selective deep price cutting by a retailer of a few well-known items, in order to attract customers to his store. Loss-leader selling is often and influentially claimed to furnish a good reason for the legalization of resale price maintenance. It is said that the price reductions involved are so limited in scope as to be of little benefit to consumers; that the practice is misleading in creating a possibly spurious impression of generally low prices; that it promotes monopoly in retailing; and that it causes meaningless fluctuations in the sales of brands selected as leader items, which disrupts their economical production. Since resale price maintenance removes the brands or products from the scope of loss-leader selling, it is said to protect both consumers and manufacturers. Precise evidence of the extent of undesirable loss-leader selling and of its alleged detrimental effects is, however, lacking or elusive. Resale price maintenance is in any event a crude method of stopping such selling: it precludes all forms of retail price competition, not only loss-leader selling, and the latter would presumably persist unless all products suitable for loss-leader tactics were price-maintained.

**Public policy.** Public policy toward resale price maintenance takes a variety of forms in different countries, as a few examples will show. It has been prohibited in Canada since 1951 as a restrictive practice against the public interest. The most common method of enforcing resale prices, the withholding of supplies from offending distributors, is prohibited (save in special circumstances specified in an amending statute of 1960). In the United States resale price maintenance is authorized in some but not all the states, and also in interstate commerce between "fair-trade" states. Differences in state laws and restrictive judicial interpretation of fair-trade laws have contributed to difficulties of enforcement and the collapse of the practice in the electrical goods and some other trades. In some states there is mandatory resale price maintenance in the liquor trade. The relatively favorable attitude toward resale price maintenance is regarded as anomalous by many observers in view of the per se illegality of horizontal price fixing, and the fact that resale price maintenance produces the main effects of horizontal price fixing in those retail trades in which it is comprehensive. In Sweden the practice has been prohibited since 1953, with provision for specific exemptions by a statutory body in cases where it is satisfied that such price main-

tenance reduces costs, benefits consumers, or otherwise promotes the public interest. The initial exemption of books is being terminated, and that of herrings has already been withdrawn.

In the United Kingdom enforcement of resale prices by manufacturers individually was facilitated by legislation until 1964, although a commonly used and powerful method of enforcement, the collective withholding of supplies by groups of manufacturers, had been prohibited in 1956. Legislation introduced in 1964 prohibits resale price maintenance, except for classes of goods specifically exempted by a special court. Exemption is granted only where the court is satisfied that the practice injures the interests of consumers or users less than it promotes them. As is the case in Canada, the legislation allows manufacturers individually to withhold supplies from dealers engaging in loss-leader selling of their goods.

Resale price maintenance is unimportant in underdeveloped countries, partly because enforcement is particularly difficult where there are several successive stages of distribution between supplier and final consumer, and also because the practice is inexpedient for manufacturers where their goods are resold in widely varying circumstances.

B. S. YAMEY

[See also INTERNAL TRADE.]

#### BIBLIOGRAPHY

- BORK, ROBERT H. 1966 The Rule of Reason and the Per Se Concept: Price Fixing and Market Division. *Yale Law Journal* 75:373-475. → Bork's views on vertical price fixing are discussed by J. R. Gould and B. S. Yamey in Volume 76 of the *Yale Law Journal*, pages 722-730.
- BOWMAN, WARD S. JR. 1955 The Prerequisites and Effects of Resale Price Maintenance. *University of Chicago Law Review* 22:825-873.
- CANADA, RESTRICTIVE TRADE PRACTICES COMMISSION 1955 *Report on an Inquiry Into Loss-leader Selling*. Ottawa: Department of Justice.
- GAMMELGAARD, SØREN 1958 *Resale Price Maintenance*. Project No. 238. Paris: European Productivity Agency of the Organisation for European Economic Co-operation.
- GOULD, J. R.; and PRESTON, L. E. 1965 Resale Price Maintenance and Retail Outlets. *Economica New Series* 32:302-312.
- GREAT BRITAIN, BOARD OF TRADE, COMMITTEE ON RESELL PRICE MAINTENANCE 1949 *Report*. Papers by Command, Cmd. 7696. London: H.M. Stationery Office.
- GRETHER, EWALD T. 1939 *Price Control Under Fair Trade Legislation*. New York: Oxford Univ. Press.
- SELIGMAN, EDWIN R. A.; and LOVE, ROBERT A. 1932 *Price Cutting and Price Maintenance: A Study in Economics*. New York and London: Harper.
- TAUSSIG, FRANK W. 1916 *Price Maintenance*. American Economic Association, *Papers and Proceedings* 28:

170-184. → Also published as a supplement to Volume 6 of the *American Economic Review*.

TELSE, LESTER 1960 Why Should Manufacturers Want Fair Trade? *Journal of Law and Economics* 3:86-105.

U.S. FEDERAL TRADE COMMISSION 1945 *Report of the Federal Trade Commission on Resale Price Maintenance*. Washington: Government Printing Office.

YAMEY, B. S. 1954 *The Economics of Resale Price Maintenance*. New York and London: Pitman.

YAMEY, B. S. (editor) 1966 *Resale Price Maintenance*. London: Weidenfeld & Nicolson; Chicago: Aldine.

## RESEARCH AND DEVELOPMENT

### I. INDUSTRIAL RESEARCH AND DEVELOPMENT

Jacob Schmookler

### II. FINANCING SOCIAL RESEARCH

Henry W. Riecken

#### I

#### INDUSTRIAL RESEARCH AND DEVELOPMENT

While, as commonly used, "research" describes the quest for new scientific or engineering knowledge and "development" refers to the creation and the reduction to practice of new products or processes, "research and development" (R & D) is less than the sum of the two, for the phrase ordinarily denotes only research and development occurring in organizations established to perform them. The word order of the phrase may convey the impression that scientific or engineering discovery invariably precedes development, but development often leads to new discovery or occurs without it. Industrial R & D, to which this article is largely confined, is that portion of all R & D performed by or for business. It thus excludes R & D by nonindustrial institutions for themselves or for each other but includes R & D by such institutions for industry.

Research and development are not only ancient human activities but continue outside as well as inside R & D establishments. Although the importance of R & D has undoubtedly increased greatly, both absolutely and relatively, in recent decades, it is noteworthy that of 61 major inventions in the twentieth century, about half were made by independent inventors. As recently as 1953, less than half of all inventions patented in the United States came from corporate R & D establishments. The scientists, engineers, and supervisors in the operating divisions of industry contributed about half as many inventions to their employers' patent portfolios as did their R & D counterparts. The greater part of the reduction in production costs over time, even in the largest enterprises, typically results from hosts of minor, often unrecorded, changes in technique, introduced at the site by production

engineers and others, not from the larger innovations stemming from the enterprises' own R & D establishments. Hence, if based solely on statistics of R & D expenditures or employment, estimates of the volume of all resources allocated to advancing science and technology are too low, whereas, because of the greater relative growth of R & D, estimates so based of the rate of growth of all resources are too high.

Thus, the distinctive feature of R & D is not in the ends sought but in the formal organization of the means of attaining them, i.e., in the employment of professional scientists and engineers by lay bodies, private or public, for the specific purpose of advancing science and technology. In modern times the state—the princes and absolute monarchs of the seventeenth and eighteenth centuries—first provided the support for such employment, initially mainly as patrons of the New Learning; later, and increasingly, in the hope that discoveries and inventions of military, industrial or medical value would result. Not until the 1860s did business enterprise begin to consider that science had matured enough to make its systematic cultivation profitable at the level of the individual firm. For various reasons, chief of which, perhaps, was the superiority there of scientific and technological education, that development took place first in Germany, in the dyestuff industry, under the stimulus of the Englishman William H. Perkin's accidental discovery of synthetic mauve. The rise of industrial R & D under the auspices of individual business enterprises followed in other countries at varying intervals, in each case roughly determined by the extent of the lag in quality and character of the requisite education.

General Electric's laboratory, established in 1901, was the first in the United States and was followed by du Pont's in 1902; but not until World War I cut off the United States and the Allies from German pharmaceuticals and dyestuffs were their chemical industries able to develop appreciably in the fields pioneered by the German industry. The roughly one hundred private R & D establishments on hand in the United States at the start of the war tripled, with government aid, during its course. Most of the new ones survived in the postwar years, and by contributing to the stream of product and process innovations they helped drive more and more rivals to establish R & D programs of their own. The experience of World War II and the Korean War duplicated, on a larger scale, that of World War I, except that the number of American firms engaged in R & D has apparently not increased appreciably since 1953, growth since then



taking the form primarily of increased R & D per firm engaged.

The strong association of industrial R & D and defense explains why the aircraft and missiles industry alone accounted for one-third and the electrical and communications equipment industry for almost one-fourth of all R & D outlays in private firms in the United States in 1960. However, when only private R & D expenditures are considered for the same year, the chemical and allied products industry ranks first, with about one-fifth of the total, followed closely by electrical and communications equipment.

The pressure of defense requirements on industrial R & D is also reflected in the fact that the four American industries which received the largest amounts of federal funds for R & D in 1960—machinery, instruments, aircraft and missiles, and electrical and communications equipment—were the only industries that year to use at least 80 per cent of the R & D funds at their disposal for development alone. The expenditures in these industries on basic research in that year amounted to only 1 per cent of total R & D expenditures in aircraft and missiles, 3 per cent in electrical and communications equipment, and 2 per cent each in machinery and instruments. By contrast, basic research expenditures amounted to 18 per cent of total R & D expenditures in petroleum refining and extraction, 12 per cent in chemicals and allied products, 7 per cent in primary metals and food, and 4 per cent in industry as a whole. The small ratio of basic research to total R & D in the defense-linked industries is believed to result both from the characteristic complexity of the equipment produced and the arduous testing required and from the pressure for immediate results which usually attends publicly financed R & D expenditures in the United States, particularly in defense fields. However, other, more complex factors may also be involved.

**Interindustry variations.** The relative importance of R & D differs greatly among industries. In terms of company-financed R & D per employee among companies engaged in R & D, the leading industries in the United States in 1960 were chemicals and allied products, with outlays of \$980 per employee, followed by the instrument industry, with \$710, and the electrical and communications equipment industry, with \$600. The average for all companies performing R & D was \$390 per employee. At the bottom were textiles and apparel, with \$60, and lumber, wood products, and furniture, with \$90. Since industries differ greatly in the proportion of companies engaged in R & D

(and in the same direction as R & D outlays per employee), even these data considerably underestimate the true interindustry differences. Similarly, among the companies engaged in R & D in 1957 and having 2,500 or more employees, company outlays on R & D per dollar of capital expenditure varied from highs of \$2.83 in scientific and mechanical measuring instruments and \$1.30 in drugs and medicines to lows of 37 cents in industrial chemicals and five cents in primary metals, with an average of 41 cents for all industries surveyed.

While such differences presumably reflect corresponding interindustry differences in such factors as market opportunity, technological potential, durability of equipment, and comparative advantage of professional over nonprofessional research and development, the relative weight of these underlying determinants remains to be established. Indeed, while a profit-maximizing firm would, under conditions of certainty, allocate funds between physical plant and R & D so as to equalize the return to the marginal dollar spent on each, the actual relationships are unknown because data needed for calculating the returns to R & D under typical circumstances are nonexistent. Similarly, efforts "to explain interindustry or interfirm differences in growth rates in terms of correlated differences in R & D rates have not been wholly successful thus far, because the lines of causality run in both directions.

**Organizational considerations.** As implied above, in Western nations generally, private firms now do most of the R & D, whether for their own account or the government's. This tendency is most pronounced in the United States, where in 1962/1963, for example, of an estimated national total of \$16,400 million spent on all R & D, 70 per cent was spent in, and 30 per cent was spent by, business firms—compared with 16 per cent spent in, and 67 per cent by, the federal government. Of the R & D expenditures of business enterprise, only 3 per cent was for work done in nonprofit institutions, and most of that was on behalf of the individual spending enterprise. By contrast, in the Commonwealth countries and western Europe relatively more government R & D is done "in house," and relatively more R & D for industry is performed in industry-operated research centers, financed by industry alone or jointly with government. Its traditionally more confined role has caused the government of the United States to eschew a major direct part in advancing the technical arts, outside of the military and health fields, and such new and expensive fields as nuclear

power and space, which are felt to be politically or militarily important, or such old, economically distressed and politically powerful fields as agriculture and coal mining. The minuscule amount of cooperative research in American industry is explained in part by the fact that the patent laws create monopolies that would be illegal under the antitrust laws if created by other means.

Debate on the relative merits of alternative ways of organizing industrial R & D, as well as on such related matters as the efficacy of existing patent laws—and particularly whether the government or the private contractor should hold title to patents on government-financed business-produced inventions—has swelled with the increase in R & D expenditures and with the growing recognition of the importance of technological progress to economic growth.

Thus, some argue that a shift in the site of industrial R & D from the individual firm to an industry center might eliminate wasteful duplication, permit economies of scale in R & D and the pursuit of projects too expensive for a single enterprise, and in principle make immediately available to all what would otherwise be patented or kept secret. There is, however, some danger that under such an arrangement unusual but occasionally great projects would be neglected; that individual firms would not risk innovating, when rivals could freely imitate after most of the risks had been eliminated by the pioneering firm; and that being removed from the point of application and lacking competitive pressure, R & D would be inefficiently conducted. Only about 15 per cent of the inventions patented by the U.S. Department of Agriculture and the Tennessee Valley Authority are used by private industry (although they were developed for that purpose), while between 50 and 60 per cent of inventions patented by private firms are used commercially. Since the R & D programs of the two agencies resemble those of industry centers, these differences in utilization rates have obvious pertinence to the question.

Undoubtedly the growth of elaborate, technologically complex, and extremely expensive systems has increased the optimal size of some R & D projects, as in the armaments, space, power, and communications fields. Hence, some individual R & D projects cost billions. This fact, also, has seemed to justify greater centralization of industrial R & D to insure efficient articulation of interdependent elements of industrial technique and apparatus. However, the weight of present evidence suggests that decision making in large R & D projects may already be too centralized in the United

States when radical advances are sought, because the discoveries and inventions on which such advances must be based cannot be forecast with the precision needed to preplan their interrelations.

**Scale considerations.** On another plane, some have urged relaxing the antitrust laws because of presumptive economies of scale in R & D. Large firms, it is argued on theoretical grounds, can afford R & D projects that small ones cannot. Large firms, having a longer time-horizon, can wait longer for returns; having better access to capital, they can more readily implement R & D findings; being more diversified, they are more likely to find useful the unexpected results of R & D; and having a more protected market position, they are able to capture more of the gains from new knowledge and so have a greater incentive to finance any given R & D project. These factors, it is argued, explain why, among American manufacturing firms in 1953, for example, only 10 per cent of those with 8 to 499 employees had or financed R & D programs, compared with 42 per cent for firms with 500 to 999 employees, 60 per cent for firms with 1,000 to 4,999 employees, and 94 per cent for firms with 5,000 or more employees; and they also help explain why in 1960 company-financed R & D for American manufacturing firms engaged in R & D amounted to 2 per cent of sales for firms with 5,000 or more employees, compared with 1.4 per cent for firms with 1,000 to 4,999 employees. The greater progressiveness of large firms, which such data and theoretical reasoning appear to imply, it is maintained, cannot be fully exploited if the individual firm is prohibited by law from achieving dominance in its market.

However, small firms appear to have certain offsetting advantages in the domain of R & D. Because the small firm has shorter lines of communication, its R & D scientists and engineers are likely to know the firm's production problems and the product requirements of its customers more precisely than are similar personnel in a large firm. Perhaps this is one reason why small firms use commercially about three-fourths of the inventions they patent, whereas big firms use only half. The greater contact of the scientists and engineers of the small firm with the problems the firm confronts may result in less lost motion in the R & D process itself, for firms with under 5,000 employees spend half as much on R & D per patented invention as do larger firms. Moreover, research and development on the part of operating scientists, engineers, and supervisors probably occurs more often in small than in large firms, since the division of labor in the latter is characteristically



narrower and more formal. Hence, variations in R & D per dollar of sales between firms of different sizes are a misleading index of the corresponding variations in the amount of new knowledge produced within the firms. Finally, the relative frequency of projects so large that only a giant enterprise can mount them does not seem high, for after a certain, not very large size, further increases in the size of the firm are not associated with any increase in R & D as a percentage of sales or with an increase in the proportion of firms engaged in R & D.

The ambiguity of existing evidence suggests that the net effect of different patterns of industrial organization and of the organization of industrial R & D upon the discovery and use of new knowledge may be small, perhaps because a strikingly favorable effect of a given pattern in one direction tends to be more or less offset by adverse effects in other directions. In general, it seems probable that an acceptable analysis of the problem must await a firmer understanding of the processes by which new knowledge is created and disseminated.

JACOB SCHMOOKLER

[See also INNOVATION and PATENTS. Also related are ENGINEERING and the articles under SCIENCE.]

#### BIBLIOGRAPHY

- BARBER, BERNARD; and HIRSCH, WALTER (editors) 1962 *The Sociology of Science*. New York: Free Press.
- CARTER, CHARLES F.; and WILLIAMS, BRUCE R. 1957 *Industry and Technical Progress*. New York: Oxford Univ. Press.
- CARTER, CHARLES F.; and WILLIAMS, BRUCE R. 1958 *Investment in Innovation*. New York: Oxford Univ. Press.
- FREEMAN, C. 1962 Research and Development: A Comparison Between British and American Industry. *National Institute Economic Review*, No. 20:21-39.
- GRILICHES, ZVI 1957 Hybrid Corn: An Exploration in the Economics of Technological Change. *Econometrica* 25:501-522.
- JEWKES, JOHN; SAWERS, DAVID; and STILLERMAN, RICHARD 1958 *The Sources of Invention*. London: Macmillan; New York: St. Martins.
- KAPP, KARL W. (1950) 1964 *Social Costs of Business Enterprise*. 2d ed., rev. New York: Asia Pub. House.
- MANSFIELD, EDWIN 1961 Technical Change and the Rate of Imitation. *Econometrica* 29:741-766.
- MANSFIELD, EDWIN 1963 Intrafirm Rates of Diffusion of an Innovation. *Review of Economics and Statistics* 45:348-359.
- MANSFIELD, EDWIN 1964 Industrial Research and Development Expenditures: Determinants, Prospects, and Relation to Size of Firm and Inventive Output. *Journal of Political Economy* 72:319-340.
- MANSFIELD, EDWIN 1965 Rates of Return From Industrial Research and Development. *American Economic Review* 55, no. 2:310-322.
- MANSFIELD, EDWIN 1967a *The Economics of Technological Change*. New York: Norton.
- MANSFIELD, EDWIN 1967b *Industrial Research and Technological Innovation*. New York: Norton.
- NATIONAL ADVANCED-TECHNOLOGY MANAGEMENT CONFERENCE, SEATTLE, 1962 1963 *Science, Technology, and Management*. Edited by Fremont Kast and James Rosenzweig. New York: McGraw-Hill.
- NELSON, RICHARD R. 1959 The Economics of Invention: A Survey of the Literature. *Journal of Business* 32: 101-127.
- NELSON, RICHARD R. 1961 Uncertainty, Learning, and the Economics of Parallel Research and Development Efforts. *Review of Economics and Statistics* 43:351-364.
- NELSON, RICHARD R.; PECK, MERTON J.; and KALACHEK, EDWARD 1967 *Technology, Economic Growth, and Public Policy*. Washington: Brookings Institution.
- ORGANIZATION FOR EUROPEAN ECONOMIC COOPERATION 1954 *The Organisation of Applied Research in Europe, the United States and Canada*. 3 vols. Paris: O.E.E.C. → Volume 1: *A Comparative Study*. Volume 2: *Applied Research in Europe*. Volume 3: *Applied Research in the United States and Canada*.
- Patent, Trademark, and Copyright Journal of Research and Education. → Published five times a year since 1957.
- SCHMOOKLER, JACOB 1959 Technological Progress and the Modern American Corporation. Pages 141-165 in Edward S. Mason (editor), *The Corporation in Modern Society*. Cambridge, Mass.: Harvard Univ. Press.
- SCHMOOKLER, JACOB 1966 *Invention and Economic Growth*. Cambridge, Mass.: Harvard Univ. Press.
- SCHUMPETER, JOSEPH A. (1942) 1950 *Capitalism, Socialism, and Democracy*. 3d ed. New York: Harper; London: Allen & Unwin. → A paperback edition was published by Harper in 1962.
- TERLECKYJ, NESTOR E. 1963 *Research and Development: Its Growth and Composition*. National Industries Conference Board Studies in Business Economics, No. 82. New York: The Board.
- U.S. NATIONAL SCIENCE FOUNDATION *Funds for Research and Development in Industry: Performance and Financing*. → Published annually since 1957. Supersedes U.S. Bureau of Labor Statistics, *Science and Engineering in American Industry, 1953/1954-1956*.
- U.S. NATIONAL SCIENCE FOUNDATION *Reviews of Data on Research and Development*. → Published monthly since 1956.
- U.S. OFFICE OF SCIENTIFIC RESEARCH AND DEVELOPMENT 1945 *Science: The Endless Frontier*. A report to the President by Vannevar Bush, director of the Office. Washington: Government Printing Office.
- UNIVERSITIES-NATIONAL BUREAU COMMITTEE FOR ECONOMIC RESEARCH 1962 *The Rate and Direction of Inventive Activity: Economic and Social Factors*. National Bureau of Economic Research, Special Conference Series, No. 13. Princeton Univ. Press.

#### II

#### FINANCING SOCIAL RESEARCH

As the social sciences have evolved from their origin in moral and political philosophy to their present state as empirically grounded observational and experimental sciences, they have not only lost some of their speculative tone but have added the costs associated with the systematic collection of data. In this respect they share some

of the financial problems of the physical and biological sciences, even though the social sciences generally do not use such extensive (and expensive) equipment. Instead, the costs of research in the social sciences consist principally of payment for the skilled (and often professional) labor of interviewers, observers, experimenters, coders, data analysts, and those who construct questionnaires and develop tests of various kinds. In recent years outlays for high-speed electronic computers and the professional and technical labor associated with their use have increased rapidly—a trend that seems likely to continue to grow as it becomes possible (and desirable) to work with larger and larger bodies of data. The travel and other expenses associated with cross-cultural or multinational research are another factor in rising costs.

For these reasons, modern social research is too costly to be borne by the individual investigator, and yet, with the exception of some kinds of applied research, it does not produce results that have sufficient private commercial value to encourage business ventures. Hence, like the physical and biological sciences, the social sciences depend upon grants and gifts from private and public sources or on contracts for research services, usually with governmental agencies. The nature of the source of support and the similarity or difference in interests and motives on the part of the grantor and the researcher make for some interesting and difficult problems. These include the possibility of conflict between the "fundamental research" interests of the social researcher and the need which the supporting organization has for "practical" results; the possibility that the sponsor will directly or indirectly try to influence the outcome of the research or the conclusions drawn from it; and the possibility that the sponsor's interests may make it risky or impossible even to inquire into certain topics. Less sinister but just as important is the fact that the grantor's interests are likely to be directed toward immediate and very contemporaneous social problems, so that a researcher with a particular topical interest may be overwhelmed with funds for a while and, later, when interest has shifted to another question, find himself without a patron in sight. Finally, there seem to be some problems that are chronically low on a society's priority list (such as the psychology of aesthetic experience) and others that usually have a high place (the analysis of economic fluctuations, for example). None of these problems is specific to the social sciences, for an instance of each can be found in one or more of

the physical and biological sciences. Some are, perhaps, exacerbated by the close connection between social research and human welfare, economic interests, political power, and ethical, moral, and judicial issues. Since the patterns of support in the United States are highly varied, and since they are *relatively* well documented, discussion in this article will be confined to the United States.

**The role of universities.** In the United States scientific research of all kinds has traditionally been carried on largely in and through universities and operating government bureaus, rather than through independent institutes or academies of science. Furthermore, until recently at least, the largest amount of scientific research was carried on in the universities, although an increasing share of the cost of research has been borne by the government.

Accordingly, among *performers* of research the most important institution is the college or university. Private universities in the United States are supported chiefly through endowments, fees from students, and rents or royalties, supplemented by gifts (from nongovernment donors) and by grants or contracts for research or services performed. Public universities obtain their funds chiefly from the legislative bodies of the individual states or cities in which they are located, although they may also have small endowments, rents, royalties, and the like, and they usually receive substantial amounts of grant or contract funds for research and services. In both private and public universities, funds for grants or contracts for social science research may come from the federal government, from industrial corporations, from private foundations, labor unions, and other nonprofit organizations, and from private individuals.

There is considerable variety in the specific arrangements for performing research in universities. Often an individual member of a faculty will, with the help of a graduate-student assistant and a clerical worker, carry out the major share of the collection and analysis of data and the preparation of a technical article on a topic in which he has a special personal interest and which he himself has chosen and defined. At the other extreme is the large, teamwork project, in which several senior investigators and many juniors, assistants, and clerks take part. Often such projects are initiated and contracted by organizations outside the university, and the interest of the faculty member is likely to be tangential, partial, or secondary. Between these extremes one can find a wide variety of intermediate arrangements.

We can distinguish, then, two dimensions: the



size of the research project (in money, time, and personnel) and the extent to which the problem originates spontaneously from the scholarly or professional interests of the senior investigator(s). These two dimensions are not precisely parallel, but they are far from orthogonal.

To a considerable extent the sources of funds are correlated with these two dimensions of research. The lone researcher with a single assistant is more often supported through the university's own research funds, when such are available; the large team working on some problem presented by an outside agency is likely to be supported through a contract with a firm or a government agency. The chief reasons for this situation are historically rooted in both the development of methods of research in social science and the emergence (especially after 1945) of new forms of support. Beginning in the classical tradition of individual scholarship, which placed reliance on documentary evidence and direct but somewhat unsystematic observation, social science research in the universities was originally the act of individuals, working slowly, independently, and on a fairly small scale. Gradually, the introduction of new techniques, especially sample surveys and experiments, together with machine processing of data and statistical techniques, enlarged the scholar's need for assistance. The private foundation's grant or fellowship for travel, study, and field work served the individual scholar well during the period when methods were developing, but the costliness of the new techniques taxed these resources and virtually overwhelmed the universities' own resources. Social science research was severely limited, chiefly by lack of funds, through the 1940s and 1950s. There are still financially imposed limits, and the social sciences do not yet enjoy the extent of support in the United States that the physical and biological sciences do. After World War II, however, the development of government financing for social science research greatly changed the situation and permitted social scientists to take on larger projects and more extensive and complex studies than had previously been possible for any except a few.

The increased interest of government bureaus in social science research is probably attributable to three things: the demonstration, during the war, of the practical contribution of social science research to selection, training, and other manpower problems; the growth of agencies in the government (from 1935 on) that were concerned with administering social and economic legislation, with the lagging realization that such administrative efforts could not succeed without research on the

problems that the agency had to handle; and the developing awareness that basic research on society and human behavior should be supported both for its own sake and in order to develop the basis for future applied work. Since the pattern of supporting research through university facilities had already been laid down in the natural sciences; since the trained personnel were available in universities (and it could be argued that sponsored research there provided opportunities for training graduate students); and finally, since there was no existing tradition or machinery for the establishment of independent research institutes or academies, the pattern of federal support through the universities grew up.

The detailed arrangements for financing government-supported projects (which form a substantial portion of all sponsored research) at universities can be exceedingly complicated, and this article will attempt to outline only the major aspects of the most common cases. As a rule, the senior investigator on a project is a regular member of the university faculty, and normally part of his salary is paid out of the university's own funds. The junior professional members of the project group are usually paid exclusively out of the grant funds, as are the clerical and other assistants. Although this pattern is in flux, it is generally true that research staffs have enlarged more rapidly than regular faculty positions, with the result that there are sometimes, in effect, two groups of equally competent social scientists at a university: one, the regular faculty, tied to the university through teaching and through tenure appointments; the other attached more uncertainly, through the research grant, with few or no teaching opportunities and with an appointment whose duration depends solely on the duration of the grant. Indeed, this development of two groups whose link with the university is different has produced many stresses and is one of the more serious problems arising from short-term financing of large-scale research.

Agencies of the federal government, operating on annual appropriations from the Congress, have tended to make short-term commitments to university researchers. Some agencies fund research on an annual basis, while others commit funds from the annual appropriations to support selected projects for as much as seven years in the future. Research funded on an annual basis suffers from uncertainty, the need to solicit additional funds at frequent intervals, and the difficulty of providing stable support for the members of the project staff.

On the other hand, the presence of outside research funds has often enabled the university to

enlarge and enrich the opportunities for instruction and practical training of graduate students, as well as to provide technical consultation and assistance to the regular faculty.

One further aspect of the financial arrangements between the universities and the sponsors of research remains to be commented upon, namely, the support of indirect costs which accompany the establishment of research projects. As long as social science research was the work of individual scholars who used only the space, light, heat, books, journals, and other facilities that the university was accustomed to provide for its staff, there seems to have been no question raised about "overhead" costs. But with the beginning of federal government *contracts* for research, which often involved expanding space, facilities, staff, and ancillary services, the situation changed. Universities began to notice that sponsored research, while offering increased research opportunities for their faculties and students, also resulted in increases in operating costs for the university. Accordingly, a system of payment for *indirect costs* was established and has been perpetuated. Private foundation grants do not always include an overhead cost for the university, while federal government grants and contracts always do. This situation arises partly out of the fact that research in the United States has not, on the whole, been conducted through separately established institutes, where the capital costs of research facilities would be reflected in direct outlays for buildings, equipment, supporting services, and maintenance.

**The role of the federal government.** In terms of funds spent, the federal government plays a distinctly secondary role in the *performance* of social science research. While the several departments of government maintain laboratories and other research units, the research efforts in these units are smaller in size than those of their counterparts in the universities. Establishments such as the Socio-Environmental Laboratory and the Adult Psychiatry Branch of the National Institute of Mental Health, the Wright Air Development Center of the Air Force, the research sections of the Bureau of the Census, the Personnel Research Branch of the Army, and the Agricultural Research Service and Agricultural Marketing Service of the Department of Agriculture are all budgeted in the departmental appropriation made annually by the Congress, are staffed by civil servants, and in general direct their research toward the practical problems which they encounter in performing their assigned missions. A unique case is the Smithsonian Institution, which is both public

and private. Established by private initiative and receiving support, in part, from private funds, it nevertheless has a charter from the Congress and enjoys some measure of federal government subsidy, which is appropriated annually. The Smithsonian's chief contributions to social science have been in anthropology and archeology.

Most of the research *performers* within the federal government devote themselves to applied social science rather than to basic research, which receives much less support. This situation arises, no doubt, from the fundamentally practical or administrative orientation of most government bureaus and reflects insufficient recognition of the present state of social science, where a great deal of basic work must be done before the development of practical applications from a scientific base is possible. Basic research usually demands smaller amounts of money than applied work does but requires consistent funding over relatively long periods of time, during which no "useful" results may emerge. It is difficult to resist the demand for "practical" information which administrators and legislators can use. One of the habitual and trying cycles in social science is that which begins with administrative optimism; is followed by an increase in support, together with an increase in demand for immediate results; and then fades into a period of administrative disillusionment, terminated by the cutting off of funds. Instability of support hampers immediate accomplishment and discourages research workers from undertaking substantial, long-range projects that might yield greater results. These considerations, it should be mentioned, have pertinence to government, as both a performer and a source of research funds, but have also afflicted other sectors, such as private foundations, on occasion.

**The role of other performers.** Universities and federal government research units account for nearly four-fifths of the funds spent for social science research in the United States. Of the remaining one-fifth, only a small amount is spent by the private foundations, which play a very minor role as *performers*; there are very few social scientists who are active in research on foundation staffs. The Russell Sage Foundation is an exceptional case, and its operations are perhaps typical of the few performers in this area: a small staff in social psychology devotes some share of its time to the pursuit of research that its members organize and that is related both to the scientific interests of the staff and to the objectives of the foundation.

A larger share of funds is disposed of by a variety of nonprofit research institutions such as



the Stanford Research Institute (SRI), the National Opinion Research Center at the University of Chicago (NORC), the RAND Corporation, the Systems Development Corporation, the Bureau of Social Science Research, the National Bureau of Economic Research, the Institute for Social Research at the University of Michigan, and the Bureau of Applied Social Research at Columbia University (BASR). These nonprofit, nongovernmental enterprises are a postwar phenomenon in the United States. Some have university connections, but as a rule they are without endowment and without students and tuition fees. These institutions have developed and flourished simply by engaging to conduct both basic and applied research on questions that are important to somebody, usually somebody outside the staff of the institute or center. They have received support from, and conducted research for, industrial corporations, labor unions, agencies of the federal government, private foundations, individuals, and colleges and universities. The pattern of financial arrangements, if it can be called a pattern, is extremely variable, ranging from the "one-client" organizations, such as the Operations Research Office, which depends almost entirely upon federal government funds, to the more variegated institutions (NORC, SRI, BASR), which accept commissions from all of the categories of "clients" enumerated above. The chief financial-administrative problem of such units is to maintain continuity of professional personnel and the scientific integrity of their work, in the absence of any continuous support from a source whose interest in social science research is not restricted to a single problem. In other words, the task of the director of such an institute or corporation is to provide continuing funds for the operation of the whole enterprise: he must not drive his research staff to distraction by frequent and arbitrary shifts in their assignments, yet he must avoid periods of low income during which valuable staff members must be let go because their salaries cannot be paid. The glory of a director of a private nonprofit institution stems from his skill in maintaining a substantial *scientific* productivity by the agency while at the same time providing the "client" with useful answers to his (not always scientific) questions.

There are a smaller number of independent organizations which operate on a profit basis. Except for this financial difference, they share many of the characteristics and problems of the independent nonprofit institution, although they do not have even a tenuous university affiliation. Such organizations are common in the physical sciences, where the pattern was established, and there are a grow-

ing number in psychology, which receive funds from both government and industry sources.

*Social Science Research Council.* The Social Science Research Council is not itself a performer of research but acts as a source of stimulation for needed research, better training for research, and support for research. A private organization of social scientists, it has developed close relations both with private foundations and with government agencies that share its interests. The governing board of the council consists of social scientists, as does its professional staff. Its committees, which are of two kinds, are composed of social scientists, largely university faculty members who serve without remuneration. Some of these committees administer programs of research-training fellowships and grants for research; others survey the status of research in selected fields, develop plans for needed research, organize conferences, and supervise a variety of programs, for which the council receives grants from private and public sources. Its continuity is assured by the fact that it has a basic endowment from private sources, although the bulk of the work accomplished by council committees and grantees is funded by short-term grants received from various sources. The governance of the council is effectively in the hands of professional social scientists.

*Sources of funds.* The matter of where the funds for social science research come from can be dealt with more briefly than can the financial complexities of performers' activities. There are two major sources of support: the federal government (chiefly the Department of Health, Education, and Welfare, the Department of Defense, the Department of Agriculture, and the National Science Foundation) and private foundations (chiefly Carnegie, Ford, Rockefeller, and Russell Sage). The principal flow of funds is from these two sources to the universities, with considerably smaller amounts going to nonprofit institutions. Private industry also supplies funds for social science research, although data on expenditures are not easy to obtain. Private industry is probably more important as a source of funds than as a performer of research, although it is difficult to estimate the amounts spent by industrial performers because of the problem of deciding which activities should be called social science research (as distinguished from routine data collection for the purpose of making, say, a marketing decision).

*Further problems.* Besides the problems already mentioned—the growth of "secondary" research staffs at universities and the lack of continuity and stability in independent research organizations—the foregoing description of the institu-

tional arrangements for financing social science research may have raised other questions, especially in the minds of those who are not entirely familiar with the operations of the system. For example, it is obvious that the federal government, as a principal source of revenue, could exert a large, perhaps an undue, influence on social science research. Under such conditions, are social scientists able to work on important problems or are they frequently distracted and diverted either to trivial questions or away from significant but controversial ones?

It is important to realize that although university-based social scientists are heavily dependent upon government sources for support of their research, it is equally true that government agencies are heavily dependent upon university-based social scientists for advice and for the professional and technical skills needed to carry out the research which government agencies want. Virtually every government agency that makes grants or contracts for social research uses advisory committees made up of faculty members of university social science departments. Many of the staff members in agencies making grants and contracts are themselves former university professors, whose understanding of the needs, requirements, and values of the social science community is therefore fairly good. To a considerable extent, then, policies governing federal government support of social science research are shaped by the values, and indeed the interests, of those who are most affected by these policies. It is true, of course, that the particular mission of the government agency will affect the sort of social science research it supports and encourages. Some agencies, however, devote all or a substantial part of their funds to the support of basic research in the social sciences, and they presumably are governed almost entirely in their project-granting decisions by the interests and capabilities of available social scientists who apply to them for assistance. At the opposite extreme, there are a very small number of research projects that are virtually ordered by agency representatives, who conceive them, initiate the negotiations for their accomplishment, and specify how they are to be done. Such highly specific task requests play a relatively small role in shaping the given field of inquiry, both because there are comparatively few such commissions and because most university and independent research agencies feel cramped by such specific requests and avoid accepting them whenever possible. Extreme financial need is usually the reason given by the research director of an independent institute which accepts such a commission from a federal agency or, for that matter, from industry

or business. Somewhat more influence derives from the extensive support that federal agencies have provided to certain fields: for example, agricultural economics and clinical psychology. Both these fields have been the objects of truly massive support for a number of years—agricultural economics since the 1920s and clinical psychology since the 1940s. The availability of funds in these two fields has undoubtedly influenced the career choices of graduate students who had no clear commitment to a particular subfield. Perhaps the strongest influence wielded by federal government agencies has been a negative one—that is, the government has exerted influence either by deciding not to support a field or by simply failing to take responsibility for the nurturance of a particular field. In this respect, both history and political science have probably experienced greater neglect from federal government agencies than any other of the social sciences. Psychology has been relatively well supported; anthropology and sociology somewhat less well. Perhaps because of a conservatism inherent in the situation, newly developed fields of inquiry, such as the administrative sciences, have often had great difficulty becoming established and securing support through federal agencies. To some extent the same point can be made with regard to particular topics or problems within a given discipline—topics that have been undersupported or wholly ignored. Until relatively recently at least, research on the social and behavioral aspects of birth control has been regarded as taboo or at least very touchy and dangerous for a federal agency. Similarly, inquiry into problems associated with racial segregation or the social changes associated with integration has had relatively little support from federal agencies. It seems probable that research on contemporary national politics has a dim future among federal agencies. This last example illustrates the point that it is not merely bureaucratic prudence that keeps agencies from encouraging research on certain controversial topics. Rather, it is quite evident that it would be difficult to achieve detachment and objectivity in judging research proposals when the research directly or indirectly affected the institution proffering the support. For that reason it may be more appropriate for private foundations to undertake the support of research on national politics, federal government operations, the evaluation of foreign policy, and other questions on which individual federal agencies are not likely to be able to maintain objectivity and detachment.

The implication should not be drawn from the preceding paragraph that government agencies are inherently more conservative in the social science



research which they finance than are private foundations. In terms of intellectual and scientific innovation, both federal agencies and private foundations have been sometimes conservative and sometimes innovative. The chief difference arises, however, in the duration of support for a "new idea." It is characteristic of private foundation funding that after having given initial impetus to some plan, the foundation will withdraw its support and turn its attention to other ventures, whereas the federal agency is typically more likely to continue extended and dependable support of a given venture for which it has taken initial responsibility. For this reason, there may be in federal operations a certain degree of inflexibility, which grows as the number and variety of programs to which the agency is committed increases. As it is difficult to bring any given activity to an end, it becomes increasingly difficult to undertake new activities unless there is a constant increase in the funding ability of the agency. This disease, however, does afflict some private agencies and should not be thought of as something unique to the public sector. Of course, although it means that government-supported social science research is somewhat less flexible in its terms of preference, it may also mean that it is more dependable and continuous.

Finally, it should be clear that the growing support for social science research at universities has had certain consequences for the style of life and pattern of organization at these institutions. Increased support has made it possible to undertake larger-scale projects, often in remote and inaccessible parts of the world, and often concerned with a much larger volume of data than social scientists of fifty years ago would have believed possible to handle. This growth in size and in sophistication has provided better opportunities for training future social scientists during their graduate school years, and it has produced new roles of an administrative, managerial, and entrepreneurial sort for faculty members. Although the day of the lone scholar working in a library or with direct personal observations has far from vanished, nevertheless, the spread of larger research teams has demanded an increase in scholars with managerial abilities. The head of a large research team must be skilled at defining the problem to be studied, at recruiting and employing personnel, and at managing the logistics of the operation. Above all, under the project system of grant and contract making, he must have the entrepreneurial skill necessary to secure, often from more than one source, the required support and to maintain the

flow of funds, so that the continuity of the team effort will be upheld.

Furthermore, the existence of team efforts of this sort has increased the pressure on university administrations to establish institutes, interdepartmental committees, and other interstitial organizational units within the traditional academic structure. Such a demand for new organizations does not reflect, of course, simply the presence of additional funding but rather indicates the existence of new strains, new lines of interest, and new intellectual developments which may seem to justify novel organizational arrangements. The role of outside funds in this kind of development is simply to make possible what might have been difficult or impossible within a traditional departmental structure. To repeat, such institutional innovations are not new (the Institute of Human Relations at Yale University was established in the 1920s), but the trend has been accelerated by increased funding in recent years.

HENRY W. RIECKEN

[See also FOUNDATIONS; UNIVERSITIES.]

#### BIBLIOGRAPHY

- ALPERT, HARRY 1959 *The Growth of Social Research in the United States*. Pages 73-86 in Daniel Lerner (editor), *The Human Meaning of the Social Sciences*. New York: Meridian.
- U.S. CONGRESS, HOUSE, COMMITTEE ON GOVERNMENT OPERATIONS 1967 *Use of Social Research in Federal Domestic Programs: Staff Study for the Research and Technical Programs Subcommittee*. 4 parts. 90th Congress, 1st Session. Washington: Government Printing Office.
- U.S. NATIONAL SCIENCE FOUNDATION *Federal Funds for Research, Development, and Other Scientific Activities*. → Published since 1950/1951. Volumes 1 to 12 were published as *Federal Funds for Science*.

## RESIDENCE

See HOUSING; KINSHIP; NEIGHBORHOOD.

## RESOURCES

See CONSERVATION; PLANNING, SOCIAL, *article* on RESOURCE PLANNING; WATER RESOURCES.

## RESPONSE SETS

It is usually presumed that the score a person obtains on a psychological test is determined by the content of the items and that this score reflects such characteristics as the examinee's knowledge on achievement tests, his abilities on aptitude tests, his interests on interest inventories, his traits on

personality measures, and his opinions on attitude scales. Evidence in support of this presumption is important in establishing the meaning of the score. The responses a person makes to a test, however, are a function not only of item content but also of the form of item used and of other aspects of the test situation. The item form or the test directions, for example, may induce temporary preferences that systematically influence responses, such as the tendency to respond quickly rather than accurately. In addition, the examinee may characteristically bring to tests of certain formats various test-taking attitudes and habits, such as the tendency to agree, that produce a cumulative effect on his score. Properties of test form, then, may differentially influence an individual's mode of item response and may also permit the operation of preferred or habitual styles of response. These stylistic response consistencies are called response sets. A response set is a habit or temporary disposition that causes a person to respond to test items differently than he would if the same content were presented in a different form (Cronbach 1946, p. 476; 1949).

A test presumed to measure one characteristic may thus also be measuring another characteristic (response set) that might not have been reflected in the score if some other form of the test had been administered. In this discussion the term "form" is used loosely to include all aspects of the test situation to which the examinee may react, including the form of the item, the tone of its statement, the number and nature of response alternatives provided, the test directions, instructions for guessing, and the presumed use to which the scores will be put.

Response set variance is of two major types: (1) transitory consistencies limited to a particular test or to a single testing session and (2) reliable, stable consistencies with some generality over tests and situations. Two forms of stable consistencies may also be distinguished in turn: (a) the type that reflects durable but relatively trivial individual differences (perhaps in language usage or expressive habits) and (b) the type that reflects significant aspects of personality (Cronbach 1949). To emphasize this latter type of stable response-set variance and its potential usefulness in the measurement of personality characteristics, Jackson and Messick (1958) suggested that it be renamed "response style." A response style, then, is a consistency in the manner of response to some aspect of test form other than *specific* item content; it is relatively enduring over time and displays some generality in responses both in other

tests and in nontest behavior (Jackson & Messick 1962, p. 134).

**Kinds of response sets.** The operation of response sets of various kinds has been detected on tests of several forms in widely different content areas, including ability and achievement tests, personality questionnaires and checklists, interest inventories, performance ratings, and attitude scales. The following are some of the major kinds of response sets investigated thus far.

**Tendency to gamble or guess.** Reliable individual differences have been noted in the tendency to guess when in doubt (see Cronbach 1946). These gambling consistencies vary over a wide range, from responding only when certain to attempting every item. Because he usually has partial knowledge, the frequent guesser tends to receive higher scores on ability tests than the more cautious respondent; and this difference in scores is not entirely eliminated by corrections for chance or random responding. Although one might attempt to assess partial knowledge and take it into account (Coombs, Milholland, & Womer 1956), a simpler approach to the reduction of gambling's extraneous influence on scores would be to use (or claim to use) scoring formulas that penalize guessing. However, even though the threat of severe penalties may reduce the incidence of guessing on the average, bold students will still tend to guess more than cautious ones, and differences in gambling propensities will still influence individual scores. Because of this, Cronbach (1950) has recommended that examinees be directed to answer every item, except on those occasions when the response set is intended to measure some personality characteristic, such as cautiousness and risk taking (Swineford 1938; Messick & Hills 1960; Kogan & Wallach 1964).

**Speed vs. accuracy.** When speed is an important aspect of test performance, as when the time limit is insufficient for the completion of the test, individuals tend to differ reliably in their preferences for responding rapidly as opposed to carefully and accurately (Cronbach 1950). This set to work rapidly may be related to the gambling tendency, whereas the opposing preference for carefulness appears to represent a more cautious test-taking strategy.

**Evasiveness, indecision, and indifference.** Substantial reliability has been repeatedly demonstrated for the tendency to use the noncommittal middle category on several response options, such as the neutral category on attitude scales, the "?" on the "yes?-no" format, the "indifferent" choice on the "like-indifferent-dislike" option, or the



"uncertain" response on "agree-uncertain-disagree" (Broen & Wirt 1958; Cronbach 1946; Lorge 1937). Possible bases for these preferences seem manifold. They may stem from a motivation to evade revealing a definite position. They may reflect indecision in the face of difficult choices or indifference in the face of uninteresting ones. As a cautious unwillingness to commit oneself, they might be inversely related to the gambling tendency discussed above. Or they may reflect consistent differences in the interpretation of category labels.

*Interpretation of judgment categories.* Reliable preferences for particular response options have been observed that appear to be partly due to stable differences in viewpoint about the meaning and scope of the judgment categories provided (Cronbach 1946; 1950). Thus, for example, two persons with the same pattern of interests would obtain quite different scores if one consistently interpreted the "like" category in a "like-indifferent-dislike" format to include anything that he did not dislike while the other limited its application to those things that he actively desired.

*Extremeness.* Another example of preferences for particular response options is the tendency to mark extreme categories as opposed to more moderate ones on rating scales and Likert-type formats (which permit degrees of agreement and disagreement). Individual differences in the use of extremes on rating scales are very reliable and, for attitude scales at least, appear to be relatively independent of item content (Kogan & Wallach 1964; Peabody 1962). In addition to its possible occurrence as a result of differences in the interpretation of judgment categories, the tendency to respond extremely may also stem from a "desire for certainty" or may reflect self-confidence in expressing opinions (Brim & Hoff 1957; Kogan & Wallach 1964).

*Confidence.* The tendency to select the option "correct answer not given" in multiple-choice ability tests has been suggested as a response-set measure of self-confidence (Mullins 1963). The tendency to choose this alternative on a verbal test was found to be significantly correlated with the same tendency on a spatial test, even with ability held constant.

*Inclusiveness.* When no specific limit is placed on the number of responses required (as in instructions to "list the activities that interest you" or "mark those statements that reflect your attitudes"), some individuals consistently tend to give many responses while others tend to give few (Broen & Wirt 1958; Cronbach 1946; 1950). The

tendency to be broadly inclusive in responding may be a reflection of individual differences in the perceived breadth of categories used in classification (Pettigrew 1958), or it may reflect uncriticalness in delimiting these categories and in establishing criteria for inclusion.

*Criticalness.* Consistent differences in the strictness of evaluating the equivalence of objects or their acceptability in terms of some standard have been interpreted as a response set of criticalness. Scores for this disposition (for example, the tendency to respond "different" in appraising the equivalence of two possibly alternative expressions or to respond "ambiguous" in judging the acceptability of a sentence as an unambiguous statement) were found to be reliable, significantly intercorrelated, and generally unrelated to content measures from the same tasks (Frederiksen & Messick 1959).

*Acquiescence.* The tendency to answer "true" on true-false examinations and tendencies to respond "agree," "like," and "true" on personality, interest, and attitude questionnaires have been found to be reliable and stable over time (Cronbach 1942; 1946; 1950; Couch & Keniston 1960; Jackson & Messick 1958). This acquiescence tendency operates primarily when the respondent lacks knowledge or certainty about his answers. On achievement tests, then, acquiescence would be expected to reduce the spread between knowledgeable and poor students on true items and to increase it on false items, thereby lowering the reliability and validity of the "trues" score and raising that of the "falses" (Cronbach 1942). On personality and attitude scales, acquiescence would tend to occur when items are ambiguous, vague, or so neutral that their desirability or undesirability is unclear (Jackson & Messick 1962). Individuals acquiescing to personality questionnaires were shown in one study to possess personality characteristics of extroversion, expressiveness, and impulsivity (Couch & Keniston 1960), thus suggesting that such acquiescence has some basis in stylistic consistencies of temperament.

There are several indications that the tendency to agree with extreme and sweeping generalizations is a different, although correlated, response style that has been called the tendency to "over-generalize" (Jackson & Messick 1958). Measures of this tendency have been found to have significant negative correlations with both criticalness and verbal ability (Frederiksen & Messick 1959), as well as with the abilities to overcome embeddedness and distraction (Forehand 1962). This suggests that the tendency to overgeneralize is more

a function of cognitive and intellectual limitations than it is of temperamental dispositions.

*Tendency to respond desirably.* The tendency to respond not to specific item content but to a more general connotation of an item's meaning—namely, its desirability—has been found to be highly reliable (Edwards 1957). Although this tendency to respond in a socially desirable way is thus not completely independent of item content, it nevertheless represents a general response consistency that is different from the more limited consistencies generated by the truthful endorsement of specific item content. It qualifies as a response set under the definition given above (Cronbach 1946), since its operation causes individuals to give different responses when the same content is presented in different forms. Thus, responses to statements in a true-false form might be determined largely by the desirability of the characteristics described; but when the same items are presented in a forced-choice form, with alternatives matched for judged desirability, the responses are more likely to be determined by specific content (Edwards 1957).

On the basis of an obtained high relation between desirability responding and content scores (Edwards 1957; Jackson & Messick 1962), a major portion of response variance on several personality questionnaires has been attributed to the set to respond desirably. The meaning of this relation, however, remains controversial. Some theorists feel that desirability responding represents an attempt to put oneself in a favorable light, while others feel that it results from an accurate description of the individual's actual desirability. It is unlikely that respondents would deliberately fabricate on such a large proportion of their responses, particularly under research conditions; but it also seems unlikely that accurate self-descriptions would produce a single pervasive dimension intimately associated with desirability that on many questionnaires overshadows the various content dimensions of personality. An alternative hypothesis interprets desirability responding not in terms of deliberate misrepresentation but chiefly in terms of an autistic bias in self-regard (Damarin & Messick 1964).

*Tendency to fake.* The tendency to fake and distort responses in an attempt to bias the impression given to the examiner represents another reliable type of desirability responding. This tendency, which must be at least partially deliberate, is usually assessed on items with a wide discrepancy between the judged desirability of the characteristic described and its frequency of occurrence.

Thus, a penchant to claim desirable but extremely rare characteristics, or to deny undesirable but extremely common frailties, would be interpreted as lying. Such items appear in several "lie" and malingering scales, which have been found to intercorrelate in factor-analytic studies to form a dimension of response that is distinct from the type of desirability responding described above (Edwards & Walsh 1964; Jackson & Messick 1962). The operation of this bias in self-report may be appraised through the use of these lie scales so that particularly suspect respondents may be detected.

*Tendency to deviate.* Consistencies have been noted in the tendency to deviate from a modal response or from the typical response of some criterion group. Since pathological groups, which are deviant from normals in critical ways, also display deviant response styles on certain tests (such as a picture preference test), the generality of deviant response patterns has been hypothesized (Berg 1955). The implications of this hypothesis and its empirical basis were critically evaluated by Sechrest and Jackson (1963), who expressed reservations about the claimed generality.

*Measurement and control of response sets.* Several different kinds of response sets have been isolated that are reliable and stable and that reflect varying degrees of generality. Their operation may either increase or reduce the reliability and empirical validity of scores. However, since they cause individuals with equal status on an ability or trait to receive different test scores, response sets seriously attenuate logical validity and complicate interpretations (Cronbach 1946). Thus, response sets tend to introduce errors of measurement that should be avoided and controlled.

Procedures for controlling response sets include (1) changing the test form to prevent their occurrence, (2) modifying directions to reduce their operation, and (3) using special response-set scores to correct for their influence (Cronbach 1950). If possible, response options with fixed categories, such as "true-false" or "yes-?-no," should be avoided in favor of multiple-choice or forced-choice forms. If fixed-response categories are used, the alternatives should be defined as clearly and as objectively as possible. Whenever the occurrence of a particular response set is thought likely, the instructions should be written to provide more structure and to reduce ambiguity. With regard to differences in guessing, for example, subjects should be specifically directed to answer all items (Cronbach 1950). In addition, special response set scores (Helmstadter



1957) may be used to correct for the set effect and to detect and discard subjects extreme in set responding.

On the other hand, some response sets appear to reflect important personality characteristics, and measures of these response styles may prove to be useful in their own right. When a response set is used as a measure of personality, the above control procedures should be avoided, and the test conditions should be designed to elicit the response set and to heighten its influence.

SAMUEL MESSICK

[See also PSYCHOMETRICS. Other relevant material may be found in ACHIEVEMENT TESTING; APTITUDE TESTING; DECISION MAKING; INTELLIGENCE AND INTELLIGENCE TESTING; PERSONALITY MEASUREMENT; SCALING; VOCATIONAL INTEREST TESTING.]

#### BIBLIOGRAPHY

- BERG, IRWIN A. 1955 Response Bias and Personality: The Deviation Hypothesis. *Journal of Psychology* 40: 60-71.
- BRIM, ORVILLE G. JR.; and HOFF, DAVID B. 1957 Individual and Situational Differences in Desire for Certainty. *Journal of Abnormal and Social Psychology* 54:225-229.
- BROEN, WILLIAM E. JR.; and WIRT, ROBERT D. 1958 Varieties of Response Sets. *Journal of Consulting Psychology* 22:237-240.
- COOMBS, CLYDE H.; MILHOLLAND, J. E.; and WOMER, F. B. 1956 The Assessment of Partial Knowledge. *Educational and Psychological Measurement* 16:13-37.
- COUCH, A.; and KENISTON, K. 1960 Yeasayers and Naysayers: Agreeing Response Set as a Personality Variable. *Journal of Abnormal and Social Psychology* 60:151-174.
- CRONBACH, LEE J. 1942 Studies of Acquiescence as a Factor in the True-False Test. *Journal of Educational Psychology* 33:401-415.
- CRONBACH, LEE J. 1946 Response Sets and Test Validity. *Educational and Psychological Measurement* 6: 475-494.
- CRONBACH, LEE J. (1949) 1960 *Essentials of Psychological Testing*. 2d ed. New York: Harper.
- CRONBACH, LEE J. 1950 Further Evidence on Response Sets and Test Design. *Educational and Psychological Measurement* 10:3-31.
- DAMARIN, FRED L.; and MESSICK, SAMUEL 1964 Response Styles as Personality Variables: Evidence from Published Multivariate Research. Unpublished manuscript, Educational Testing Service, Research Bulletin, Princeton, N.J.
- EDWARDS, ALLEN L. 1957 *The Social Desirability Variable in Personality Assessment and Research*. New York: Dryden.
- EDWARDS, ALLEN L.; and WALSH, JAMES A. 1964 Response Sets in Standard and Experimental Personality Scales. *American Educational Research Journal* 1: 52-61.
- FOREHAND, GARLIE A. 1962 Relationships Among Response Sets and Cognitive Behaviors. *Educational and Psychological Measurement* 22:287-302.

- FREDERIKSEN, NORMAN; and MESSICK, SAMUEL 1959 Response Set as a Measure of Personality. *Educational and Psychological Measurement* 19:137-157.
- HELMSTADTER, GERALD C. 1957 Procedures for Obtaining Separate Set and Content Components of a Test Score. *Psychometrika* 22:381-393.
- JACKSON, DOUGLAS N.; and MESSICK, SAMUEL 1958 Content and Style in Personality Assessment. *Psychological Bulletin* 55:243-252.
- JACKSON, DOUGLAS N.; and MESSICK, SAMUEL 1962 Response Styles and the Assessment of Psychopathology. Pages 129-155 in Samuel Messick and John Ross (editors), *Measurement in Personality and Cognition*. New York: Wiley.
- KOGAN, NATHAN; and WALLACH, MICHAEL A. 1964 *Risk Taking: A Study in Cognition and Personality*. New York: Holt.
- LORGE, IRVING 1937 Gen-like: Halo or Reality? *Psychological Bulletin* 34:545-546.
- MESSICK, SAMUEL; and HILLS, JOHN R. 1960 Objective Measurement of Personality: Cautiousness and Intolerance of Ambiguity. *Educational and Psychological Measurement* 20:685-698.
- MULLINS, CECIL J. 1963 Self-confidence as a Response Set. *Journal of Applied Psychology* 47:156-157.
- PEABODY, DEAN 1962 Two Components in Bipolar Scales: Direction and Extremeness. *Psychological Review* 69:65-73.
- PETTIGREW, THOMAS F. 1958 The Measurement and Correlates of Category Width as a Cognitive Variable. *Journal of Personality* 26:532-544.
- SECHREST, LEE; and JACKSON, DOUGLAS N. 1963 Deviant Response Tendencies: Their Measurement and Interpretation. *Educational and Psychological Measurement* 23:33-53.
- SWINEFORD, FRANCES 1938 The Measurement of a Personality Trait. *Journal of Educational Psychology* 29:295-300.

## RESPONSE SURFACES

See under EXPERIMENTAL DESIGN.

## RESPONSIBILITY

As a philosophical concept, responsibility is a correlate of freedom; as a political concept, it is a correlate of constitutionalism. Philosophically, the opposite of responsibility is external or internal compulsion; in political terms, it is arbitrariness.

It is necessary to distinguish between moral, political, and legal responsibility. *Moral responsibility* may be related only to consciousness; its extreme expression is Dostoevski's statement: "Everybody is guilty of everything." A clear example of such responsibility is the Stuttgart Declaration of the Evangelical church in Germany. Although no one blamed it for its attitude under the Nazi regime, the church made the following statement: "During long years we fought in the name of Jesus Christ against the devilish spirit of

which the terrorist regime of National Socialism was a terrible expression. Nevertheless we accuse ourselves that we did not confess more bravely, pray more faithfully, believe more joyfully, love more intensely." Moral responsibility may also have an objective dimension. According to John Stuart Mill, "A government is to be judged by its action upon men, and by its action upon things; by what it makes of the citizens, and what it does with them; its tendency to improve or deteriorate the people themselves, and the goodness or badness of the work it performs for them, and by means of them" ([1861] 1958, p. 28). Here moral and political responsibility are linked.

*Political responsibility* is more rigid than moral responsibility because it is judged by results and not by intentions. Political responsibility means the right use of power. It is relevant to both the tiny fragment of power in the hands of a single voter and the unlimited power of the dictator.

*Legal responsibility* is a fragile instrument of control because legislators and judges can easily misuse their political power by making laws and decisions weighted against political opponents. The ancient use of impeachment suggests how easily politics can be turned into law and law into politics.

The classic political expression of responsibility is found in Constant's *De la responsabilité des ministres* (1815). Although Constant's reasons for writing this book are now only of historical interest (the solemn impeachment power of parliament in cases of treason and embezzlement), he did develop three thoughts that still have fundamental meaning: (1) He replaced the notions of treason and embezzlement by the more general term "misuse of power." (2) He declared that in a constitutional monarchy a vote of no confidence by parliament must inevitably lead to the resignation of ministers. (3) He asserted the coresponsibility of lower executive officials who obey illegal orders.

### Fundamental concepts

The three statements above foreshadow the four main categories used in this article in the analysis of responsibility.

**Who is responsible?** Every member of the body politic is responsible according to his function, from the absolute ruler to the common voter. There are specific forms of political and legal responsibility attached to specific functions, e.g., to those of ministers and officials in constitutional states.

**How far does responsibility extend?** In the framework of a bureaucratic hierarchy, responsibility can be determined according to duties. The minister is responsible for his department, the

mayor for his city, the official for his office. The lower the position, the easier it is to define the limits of responsibility; until, at the lowest level, responsibility is limited to the execution of orders and directions. The higher the level of government, the harder it is to distinguish between the execution of laws and the exercise of discretion. In the jurisdiction of the German administrative courts the question of discretion is of great importance. After the excesses of administrative arbitrariness under National Socialism, an attempt was made to keep the field of discretion as narrow as possible. However, some discretion is necessary because efficient administration requires that each case be treated individually. Political responsibility in the broadest sense means responsibility to the public interest. The highest authorities—the sovereign people, the chief of government, the head of the state—are responsible not only for matters within their control but sometimes also for matters beyond their control, such as success or failure in war or in diplomatic negotiations. In former times, kings were held responsible even for starvation and floods. King Frederick the Great dismissed generals who were unsuccessful in battle. Public opinion tends to blame a political leader for not having mastered overwhelming odds. For instance, German political opinion held Chancellor Brüning responsible for the failure of the Weimar Republic.

**To whom is a man responsible?** Theoretically, the politician or administrator is responsible to those who gave him the mandate. His responsibility is in the first place to his superior in office and to the competent judge, who act as representatives of the sovereign; then to the sovereign; and finally to God. When the sovereign is the people, its judgment is expressed through public opinion, elections, or special tribunals. Often in primitive societies, the ultimate test of political responsibility is the ordeal.

**To which sanctions must the responsible person answer?** Legal penalties are adequate sanctions only if responsibility can be defined in legal terms. The politically responsible person can be dismissed only if he fails or has lost the trust of the person who commissioned him. The classic example is the dismissal of a minister after a vote of no confidence in parliament. Apart from this there are cases in which neither the field of competence nor the form of sanctions is fixed, but punishment is demanded by the disgraced monarch or the indignant people. Under these conditions sanctions may take the form of the "silk string" the despot sends to his vizier, of judgment by revolutionary tribunals, or of punitive measures by victorious enemies.



### Historical survey

One can speak of political responsibility only if there is a generally acknowledged standard for judging actions that affect the community; thus, political responsibility presupposes a kind of public interest. The history of political responsibility is simultaneously a history of the interpretation of public interest. In societies where a religious spirit dominates, public interest means adaptation to a divine order.

The emperor of China, for example, had to make certain that his empire found the middle path between heaven and earth. If the empire were permanently struck by misfortune, the emperor was considered to have lost his mandate from heaven; he was no longer a true emperor and could be—even had to be—deposed. Snorri Sturluson told of two Swedish priest-kings who were sacrificed to the gods in order to prevent starvation. A similar feeling for responsibility to fate was behind the Athenian institution of *ostrakismos*; a person could be banished arbitrarily, without imputation of dishonor, by vote of the majority. This was not meant as punishment for a crime, but merely to prevent some future misfortune that his continued presence might cause.

Medieval Christianity was the first to distinguish between several kinds of responsibility and between different authorities to whom responsibility was due. The church controlled the responsibility of the monarch to God. But it was a twofold responsibility. The prince had to lead a life beyond reproach in order to conform to the old idea that he personally represented the harmony of the political and the cosmic orders. At the same time he bore another responsibility, arising out of his political function: he had to dispense justice. The true nature of justice had to be determined by the public interest, which was conceived of as the best order for all or as a guaranteed order based on valid law. The absolute monarch was responsible only to God. Therefore, he was independent of the church as well as of the privileged estates. The doctrine of the "divine right of kings" included supremacy over the church. In the period of the Enlightenment the king himself separated the office from his person. Like Frederick the Great of Prussia, he thought of himself as the *premier serviteur de l'état*. The theory of the divine right of kings was upheld, but God was replaced by reason and justice as the recipient of responsibility. The king should be a *Staatskönig*; his authority was based on the idea of the state, but not necessarily on the consent of the nation.

Further changes in the concept of responsibility were introduced in the period of constitutional and parliamentary monarchies. Indeed, in these two systems the king was not responsible to anyone. In a constitutional monarchy such as Germany the ministers were responsible to parliament only for the legality of their actions. They did not need the confidence of parliament to stay in power; they needed only the confidence of the monarch. In England, a parliamentary monarchy, the ultimate decision on political responsibility rested in society itself, as represented by a full parliamentary convention.

The history of the modern state is also the history of the idea of public interest, proceeding from the privileges of the kings and the estates all the way to the *raison d'état* of absolutism. In continental Europe the *raison d'état* was generally embodied in the prince and his administrative and military staffs, and the public interest was represented by a well-ordered administration and well-ordered courts. In England the character of parliament changed; from representing the special interests of the different estates it came to represent the whole nation, and freedom became the standard of public interest.

The principle of popular sovereignty, through either representative or plebiscitarian democracy, has now been accepted all over the world. The constitutions of representative democracies contain a web of responsibility on different levels and a system of corresponding controls. Courts of justice examine whether legislation is within the limits of constitutional law and whether administration is exercised within prescribed bounds. Parliament controls the politics of the cabinet. In the same way it also controls the conduct of every minister responsible for a department. Public opinion controls the political style. Its organs are a free press, corporations of public or private law, and interest groups, spontaneous assemblies, or action groups. There is no immanent contradiction between representative and plebiscitarian democracy. But as soon as the plebiscitarian principle itself becomes an important factor, differentiation of responsibility is no longer possible. For in this case rulers and nation are declared to be identical, and the people itself takes over responsibility. Totalitarian regimes, therefore, often insist on being regularly legitimated by genuine or faked plebiscites.

### Types of responsibility and control

A distinction must now be made between the responsibility of a trustee and the responsibility of an agent. [See REPRESENTATION.] The trustee's

responsibility is personal; the agent's responsibility is technical. This distinction permits a consistent scheme for the different examples discussed. The leader of the state, whether a single person or a group, acts in the name of the whole body. Leaders can be regarded only as trustees of the whole—in a democratic society, as trustees of the whole nation. On the other hand, people responsible only for clearly circumscribed functions are mere agents. Their responsibility is limited by their instructions, provided that these instructions are in accordance with the constitution. A third and very important group consists of the leaders of the constituent sections or departments of the total power, for instance, members of legislative bodies, supreme courts, ministers as heads of the administration. In some of their duties they can be regarded as agents, but because they are more responsible for the spirit than for the manner of executive power, they are trustees. One might say that they exercise secondary trusteeship. In a democracy the distinction between trustee and agent vanishes. Even agents as a part of the sovereign people are regarded as—and may feel themselves to be—trustees. On the other hand, an authoritarian, and above all a totalitarian, government destroys any sense of trusteeship. This is illustrated by the attitudes of Hitler's and Stalin's highest functionaries, who hid behind orders and denied any personal responsibility, claiming to be only agents.

**Primary trustees.** The sovereign is responsible for the state as a whole. In democracies that responsibility is exercised through elections or plebiscites. There is no appeal against the decisions of the sovereign. It is a mutual responsibility of the people and their representatives; its standard is the spirit of the community. But there are exceptional cases, where responsibility is exercised by self-nominated agents of public opinion. Such exceptions may begin in rather trifling expressions of public uneasiness, as in the *Der Spiegel* affair in Germany in 1962. They can lead to a total change in the regime, as did *l'affaire Dreyfus* in France, and ultimately to a violent revolution, such as that caused by the protest of the French press against the royal ordinances in 1830.

The leader of the government, the head of the state, or the prime minister is also responsible for the whole destiny of a nation. He is legally controlled by parliament, by the voters, and indirectly by public opinion. He will be held responsible not only for misconduct, but also for bad policies, and in the last analysis for success or failure. The normal sanction is removal at the next election. Although other methods for removal are

available, their utility is doubtful; in the United States, for example, the institution of impeachment has never led to a conviction. The parliamentary committees for investigation, developed above all in German public law, may help to detect corruption, but they have no serious political effects. Decisions by supreme courts, e.g., the very influential Bundesverfassungsgericht (federal constitutional court) in Germany, are useful correctives for legislation and administration, but they do not provide sanctions against irresponsible behavior.

**Secondary trustees.** The importance of secondary trusteeship depends on the character of the government. It is considerable in a government with collective responsibility; it diminishes as a government approaches monarchy. Even under the same constitution it changes with the personal influence of the head of government and the minister concerned. In England the minister, as head of a department, is regarded as responsible for the conduct of his officials. Parliament will not remove him for minor blunders but probably will for mistakes seriously affecting men or institutions. In most cases the minister will resign quickly to save the government. In the United States the presidential system gives no opportunity to overthrow the government; the responsibilities of the departments and of their chiefs are sometimes not very obvious. So Congress uses other methods to get rid of unpopular officials: refusing to allocate funds, delaying bills, and so on. In the Continental systems the power of the bureaucracy as a semiautonomous body is so great, and is organized in such a way, that the main responsibility lies not with the parliamentary ministers but with the permanent secretaries of state or their equivalents. Here the special administrative courts are important means for compelling the administration to observe the intentions of the legislature.

**Agents.** The responsibility of agents consists in implementing constitutional provisions, making independent decisions of minor importance, and executing orders given by superiors. In the first two cases, the public is interested in internal control through the bureaucratic hierarchy itself and through administrative courts so that authority is not capriciously misused. The means of internal control are described by Carl J. Friedrich (1937, chapter 19) as (1) disciplinary measures, (2) promotional measures, (3) financial measures, (4) judicial measures, and (5) the spirit of craftsmanship. Friedrich makes the interesting observation that the appeal to *esprit de corps* is in some respects akin to the appeal to religious responsibility and affects the deepest strata of personality.



## Obedience and responsibility

The execution of orders given by superiors can give rise to serious conflicts. The official may be anxious to avoid a wrong policy because, as an expert, he can foresee serious consequences. In such cases the official should, according to Max Weber, "execute conscientiously the order of the superior authorities, exactly as if the order agreed with his own conviction. This holds even if the order appears wrong to him and if, despite the civil servant's remonstrances, the authority insists on the order" ([1919] 1946, p. 95). Responsibility in these instances is necessarily limited, and the official will resign only when he is convinced that the execution of the order would be dangerous for the state or would dishonor him.

Another source of conflict is the illegality of orders. In this case the official is not allowed to obey. When this occurs in a constitutional regime, it will be solved very quickly by appeals to parliament or by other legal means. The guilty superior will be reprimanded, and the disobedient official will be cleared. In authoritarian or totalitarian states the problem is more difficult. We still see Nazi officials trying to exculpate themselves with the claim of *Befehlsnotstand* (danger to their lives if they did not obey). Although the constitution and the special laws, and even military law, clearly forbid the execution of illegal orders, the culprits often justify themselves on the ground of imminent danger to their security and life. But personal responsibility must prevail even against obedience to political superiors. It is an important step in the democratic development of a people when the citizen's responsibility toward law takes precedence over his responsibility toward a superior.

O. H. v. D. GABLENTZ

[See also AUTHORITY; CONSTITUTIONS AND CONSTITUTIONALISM; DICTATORSHIP; LEGITIMACY; TOTALITARIANISM.]

## BIBLIOGRAPHY

The most complete discussion of political responsibility is in Friedrich 1937. For the American point of view, the writings of Niebuhr 1956 and Lippmann 1955 should be consulted. Fauconnet 1920 deals with criminal responsibility in different societies. Many references may be found in "The Quest for Responsibility" in Schubert 1961.

CONSTANT DE REBECQUE, HENRI BENJAMIN 1815 *De la responsabilité des ministres*. Paris: Nicolle.

FAUCONNET, PAUL 1920 *La responsabilité: Étude de sociologie*. Paris: Alcan.

FRAENKEL, ERNST 1958 *Die repräsentative und die plebiszitäre Komponente im demokratischen Verfassungsstaat*. Tübingen (Germany): Mohr.

FRIEDRICH, CARL J. (1937) 1950 *Constitutional Government and Democracy: Theory and Practice in Europe*

and America. Rev. ed. Boston: Ginn. → Originally published as *Constitutional Government and Politics: Nature and Development*.

GABLENTZ, OTTO HEINRICH VON DER 1959 *Politische Gestaltung*. Cologne (Germany): Westdeutscher Verlag.

HENNIS, WILHELM 1962 *Amtsgedanke und Demokratie-begriffe*. In Konrad Hesse, Siegfried Reicke, and Ulrich Scheuner (editors), *Staatsverfassung und Kirchenordnung: Festgabe für Rudolf Smend zum 80. Geburtstag am 15. Januar 1962*. Tübingen (Germany): Mohr.

LEIBHOLZ, GERHARD 1958 *Strukturprobleme der modernen Demokratie*. Karlsruhe (Germany): Müller.

LIPPMANN, WALTER 1955 *Essays in the Public Philosophy*. Boston: Little.

MEINECKE, FRIEDRICH (1924) 1957 *Machiavellism: The Doctrine of Raison d'État and Its Place in Modern History*. New Haven: Yale Univ. Press. → First published as *Die Idee der Staatsräson in der neueren Geschichte*. Contains a general introduction to Friedrich Meinecke's work by W. Stark.

MILL, JOHN STUART (1861) 1958 *Considerations on Representative Government*. New York: Liberal Arts.

NIEBUHR, REINHOLD 1956 *Reinhold Niebuhr, His Religious, Social, and Political Thought*. Edited by C. W. Kegley and R. W. Bretall. New York: Macmillan.

SCHUBERT, GLENDON A. 1961 *The Public Interest: A Critique of the Theory of a Political Concept*. Glencoe, Ill.: Free Press.

STAHL, FRIEDRICH JULIUS (1830-1837) 1854-1856 *Die Philosophie des Rechts*. 2 vols., 3d ed. Tübingen (Germany): Mohr.

WEBER, MAX (1919) 1946 *Politics as a Vocation*. Pages 77-128 in Max Weber, *From Max Weber: Essays in Sociology*. Translated and edited by H. H. Gerth and C. W. Mills. New York: Oxford Univ. Press.

## RETAILING

See INTERNAL TRADE.

## RETALIATION

See SANCTIONS, INTERNATIONAL.

## RETARDATION, MENTAL

See MENTAL RETARDATION.

## RETENTION

See LEARNING; FORGETTING.

## RETIREMENT

See AGING.

## REVEALED PREFERENCE

See UTILITY.

## REVENUE

See TAXATION.

## REVISIONISM

See MARXISM; SOCIALISM.

## REVIVALISM

See NATIVISM AND REVIVALISM; SECTS AND CULTS.

## REVOLUTION

Revolution in its most common sense is an attempt to make a radical change in the system of government. This often involves the infringement of prevailing constitutional arrangements and the use of force. "Revolution" may also mean any fundamentally new development in the economy, culture, or social fabric—that is, in practically any field of human endeavor. In its modern (political) sense, the term was first used in the Italian city-states in the late Middle Ages, when it referred mainly to ecclesiastical reforms. The word first entered the English language in about 1600; under Cromwell, paradoxically, it came to mean the restoration of the old order. Traditional Spanish or Latin American *pronunciamentos*, Middle Eastern *coups d'état*, and the frequent upheavals in which one ruling clique replaces another, merely substituting one king, colonel, or courtier for another, without otherwise affecting the system of government or the fabric of society, cannot be considered revolutions unless, as has happened in a few cases, such upheavals are transformed into movements that bring about radical social change. On the other hand, it would be misleading to confine examination of the subject to the French and Russian revolutions, regarding all others as merely precursors of, or deviations from, these "classic" types.

Modern revolutions are usually carried out, according to their leaders, on behalf of the popular forces against despotism, corruption, and an outworn social and political order, and under the banner of progress, freedom, and social justice. Social movements, however, have to be judged not only by intentions, ideological declarations, and promises, but also by their actual performance. Some of those who take revolutionary action in the name of freedom and social justice are demagogues or impostors; others sincerely believe in these ideals but as a result of their actions may bring about a state of affairs that is a complete negation of their beliefs. Thus it is not always easy to distinguish between revolution and counterrevolution. The rise of totalitarianism has further obscured what was in the nineteenth century a seemingly clear-cut conflict between left and right. Condorcet's dictum that only movements striving for freedom can be truly revolutionary has become less and less tenable.

**The causes of revolution.** Despite the different patterns revolutions have followed, it is possible to trace certain features common to all. They usually follow the complete or partial breakdown of the old

order caused by the inefficiency of the governing class or by economic crisis, war, or similar happenings. Yet a "revolutionary situation" per se does not necessarily result in revolution; it may just as easily lead to anarchy or to a nonrevolutionary dictatorship if the revolutionaries (usually a conspiratorial group better organized than their rivals) are not capable of swift and decisive action.

The incident that sparks a revolution may be trivial: the performance of an opera of Daniel Auber's in Brussels in 1830; a student's cane dropping into the orchestra pit from a balcony in Munich in 1848; or the punishment by bastinado of a group of merchants in Tehran in 1905. The deeper causes, of course, go further back in time, and are more complex. Temporary economic crises have played a certain part in the outbreak of revolutions (the French famine of 1788, the bread riots in Petrograd before the February Revolution in 1917). However, left-wing French historians (Jean Jaurès, Georges Lefebvre) have doubted whether the economic crisis alone would have induced the masses to participate actively in the revolution of 1789, and communist historians, commenting on the year 1917, have attributed more importance to antiwar feeling than to food shortages.

War appears to have been the decisive factor in the emergence of revolutionary situations in modern times; most modern revolutions, both successful and abortive, have followed in the wake of war (the Paris Commune of 1871, the Russian revolution of 1905, the various revolutions after the two world wars, including the Chinese revolutions). These have occurred not only in the countries that suffered defeat. The general dislocation caused by war, the material losses and human sacrifices, create a climate conducive to radical change. A large section of the population has been armed; human life seems considerably less valuable than in peacetime. In a defeated country authority tends to disintegrate, and acute social dissatisfaction receives additional impetus from a sense of wounded national prestige (the Young Turks in 1908, Naguib and Nasser in 1952). The old leadership is discredited by defeat, and the appeal for radical social change and national reassertion thus falls on fertile ground.

Established elites and ruling classes have, however, also fallen in times of peace, for such reasons as sheer ineffectiveness, corruption, general boredom, or inability to adapt themselves to changing conditions. Corruption or extreme conservatism by itself, however, does not produce revolution. Despotism is often overthrown, not when they are being most tyrannical, but rather when



the established elite loses its self-confidence and starts making halfhearted attempts to become more liberal. The French Revolution of 1789 broke out after the States-General had been called; the revolution of 1848, after the unpopular Guizot government had fallen; the Hungarian revolution of 1956, after the surviving victims of extreme tyranny had been released. A united elite with a firm belief in the rightness of its cause has hardly ever been overthrown, regardless of the extent of its political, military, or economic setbacks. On the other hand, a comparatively minor reverse may prove fatal to an otherwise successful and reasonably competent elite that has lost its self-confidence and its will and ambition to rule.

A frequent source of revolutionary tension is the social unrest that may occur when a section of the population or a social class does not receive the political rights which it feels correspond to its value in society, when it faces a rigid structure or elite, or when its social and economic demands are not fulfilled. Oligarchies tend to defend their own privileges and to ignore the demands of a newly emergent sector or class, and as a result the "outsiders" despair of the possibility of gradual, peaceful change and conclude that only the complete overthrow of the old order will give them their rights and restore justice. In this context all revolutions, from the *secessio plebis* and Marius in ancient Rome to the French and Russian revolutions, are simply the culmination of long-drawn-out social processes; not the result of some sudden dissatisfaction of part of the population, but the last phase of a process of long duration.

In a similar way the rebellion of colonies against the metropolitan country and their eventual secession is a revolutionary process, made possible by the social, economic, and political development of the colony and the gradual evolution of a local elite.

Revolutions have a tendency to spread. Events in 1848 and 1918-1919 provide the best-known examples, but this generalization is also true of comparatively minor revolutionary uprisings. The Spanish revolution of 1820 (Rafael del Riego) indirectly provoked uprisings in Portugal (Bernardo C. Sepulveda), Naples (Guglielmo Pepe), and Piedmont. The French revolution of July 1830 had repercussions in Belgium and Poland. Many Asian and African countries gained independence after India had won hers in 1947. A revolution in one country frequently acts as a stimulus to revolutionary activity in another.

**The phases of revolution.** The stages of a modern revolution vary according to time, place, and the character and aims of the revolutionary group.

The activities of a group that has set itself limited aims, as in Mexico in 1910, differ from those of a party that stands for the total transformation of society, as in China in 1949. There are, however, certain features common to all but a very few revolutions.

*Prerevolutionary outbreaks.* Few revolutions come as a bolt from the blue: demonstrations, strikes, meetings, outbreaks of violence, a partial or total breakdown of law and order usually foreshadow the shape of things to come.

*Initial moderate stage.* After it has come to power, a revolutionary movement, being a coalition of various groups and parties, often assumes a moderate character. The exceptions are revolutionary movements that have come to power after a prolonged civil war and have decimated their enemies and rivals as a result of that war (e.g., Yugoslavia in 1945, China in 1949). Once the leadership of the former regime has been overthrown and the outstanding symbols of its rule removed, the new revolutionary regime is interested above all in "business as usual," in other words, the restarting of essential services, work in factories, shops, and on the land. A total breakdown would not be in the interest of the revolution, which usually lacks the qualified manpower to fill all key positions with its own candidates and therefore needs the cooperation of a substantial number of supporters of the old regime. Some genuine revolutionary movements fail at the first stage; Chu Yüan-chang came to power in 1368 as an agrarian revolutionary, yet could not carry out any radical reforms because he needed the assistance of the wealthy landowners to pay his army. Batista's Cuban uprising in 1933 differed from the traditional Latin American *coup d'état* and had the makings of a genuine revolution, since it placed sergeants over officers; yet its sole achievement was to enable the newcomers to share the spoils of the old system.

*Radical phase.* Revolutionary movements that outlast the first phase—both those which have more ambitious plans and those which are driven forward by the sheer logic of events—tend to adopt more radical measures in the second phase. At the same time the coalition of parties usually splits up, and the more radical elements (Jacobins, Bolsheviks) emerge as the sole possessors of power. Within the ruling group power passes into the hands of a very small number, often into the hands of a single individual. This process may be hastened by counterrevolutionary activities and/or foreign intervention, but it may also happen irrespective of such opposition. Every major revolution has destroyed the state apparatus it found and eventually

replaced it by setting up another, generally stronger bureaucratic organization in its place. This centralizing dictatorial trend has been marked in every modern revolution.

In many revolutions there are two distinct phases: the overthrow of the upper aristocracy in the English Civil War, for example, was followed at a later stage by the emergence of the middle and lower classes. The February Revolution of 1917 was followed by the Bolshevik seizure of power. Once the main revolutionary measures have been carried out and a certain normality re-established, there is frequently a change of guard. Some of the revolutionary heroes and leaders succeed in adjusting themselves to the new administrative and organizational tasks facing them; others do not and are cast aside, giving way to managers and bureaucrats. In the struggle for power that often ensues, it is by no means always the most radical leader who triumphs, but the better tactician, the man with the most support within the ruling group, the army, the police, or other power base.

*Postrevolutionary situation.* How do revolutions end? Some revolutionary movements fail at a very early stage or are beaten by the forces raised against them. Some of those that last longer peter out because a key leader dies (Oliver Cromwell) or because internecine strife undermines the revolutionary party (e.g., France in 1794). Others evolve over a long period of time and gradually change their character. Certain revolutionary achievements are maintained; others are slowly eroded, although the phraseology of the revolution is usually preserved. There are unlimited variations on this theme.

*The personnel of revolutions.* The typical nineteenth-century or twentieth-century revolutionary is of middle-class or lower-middle-class origin—student, young lawyer, or junior army officer. Elderly and very rich people or those with conservative inclinations are unlikely to lead or engage in revolutionary activity. There are, as usual, exceptions: the Brazilian revolution of 1889, which led to the downfall of the monarchy, was a movement of protest by leading *hacendados* against the abolition of slavery. The railroad disturbances in China in 1911, which led to the deposition of the Manchu dynasty, began as a protest of the principal railroad shareholders against nationalization. On the whole a revolutionary mentality is unlikely to flourish in such circles, although a number of aristocrats were on the side of the *tiers état* in 1789. The majority of leaders of socialist movements and of proletarian revolutions have been of middle-class origin; this fact suggests that the

feeling of frustration, the quest for power, the sense of injustice, and various idealistic aspirations are of greater importance in the formation of a revolutionary than is economic discontent. Oppressed nationality is occasionally an important factor (the high percentage of Jews, Latvians, Armenians, etc., in the initial stage of the Russian Revolution, the prominent part played by national minorities in revolutionary movements in the Balkans and the Middle East, the role of Poland in 1848, and so on).

In those parts of the world where grave social and political problems exist, higher education acts as a powerful stimulus toward revolution. The illiterate masses may be quiescent, however miserable their lot, but higher education provokes a "revolution of expectations" that cannot be fulfilled. Students and high school pupils have played a very prominent part in both the Russian and Chinese revolutions and in revolutionary movements in Latin America, the Middle East, and other parts of Asia and Africa; the fact that these were or are countries with a high proportion of unemployment or misemployment of university graduates is not accidental. The conflict of generations is another factor of considerable importance (the revolution of the *tenentes* in Brazil in 1939; of young army officers in Venezuela in 1945, in Egypt in 1952, and in Iraq in 1959).

Some revolutionary movements have been led and supported by one specific class or group (for instance, peasant revolts), but the major revolutions have had a broader basis both as to leadership and rank and file. Intellectuals have played a leading part in most revolutionary movements in Europe since the eighteenth century, while in many countries outside Europe young army officers have had a conspicuous role. The army as such in these countries is no more revolutionary in spirit than the police or the civil service, who on the whole play a passive part (or a waiting game) and regain importance only after the victory of the revolution; but as the principal source of physical power, the army is the most obviously effective instrument of revolution. Young revolutionaries in the Middle East and in Latin America have often chosen to become army officers because this career was the most likely to give them power and thus bring about political and social change.

Since, as a rule, a revolution can succeed only if preparations for it are kept secret, these have usually been confined to a small revolutionary high command. In the era before modern political parties came into being, an important part was played by secret societies, such as the clubs in France on



the eve of 1789 or the Chinese secret societies, which traditionally had great political influence (the White Lotus in 1775, the Society of Heaven's Law in 1813). A certain amount of conspiracy is involved in every revolution, but its ramifications are often exaggerated (the "hidden hand," the "conspiracy theory" of history) and undue importance subsequently attributed to certain bodies (the Illuminati, the Freemasons), attributing causal or organizational connection to events that were in fact quite unconnected. Apart from what is planned, there is in each revolution a spontaneous element, which is sometimes far more decisive. The revolutions in Germany and Austria in 1918 were neither foreseen nor planned by any revolutionary high command.

Revolutionary uprisings undertaken by the army are usually prepared by a single commander or a small group of officers (*junta*) having the support of a substantial part of the army. The garrison of the capital or the units stationed nearby are usually of decisive importance in this context. Twentieth-century revolutions in the more highly developed countries need a mass basis; the army may be essential for the initial stage, but different instruments are required to hold power and to carry out the changes deemed necessary. The more ambitious military usurpers have therefore tried either to collaborate with some of the existing political parties or groups or to create a political mass movement of their own (Perón, Nasser, etc.).

**Success or failure of revolution.** The factors determining success or failure in revolution are as various as the factors that generate revolutionary activity in the first place. The most frequent causes of failure are lack of support or active resistance by the bulk of the population and disunity or lack of purpose on the part of the leaders. In all modern revolutions the support of a militant minority and the physical seizure of some vital *points d'appui* (such as the seat of government, the army and police commands, the means of mass communication—radio, press, etc.) are decisive. These steps are carried out by a comparatively small minority and stand a good chance of succeeding if the group has mastered the elementary technique of the *coup d'état*, if the enemy is disorganized and incapable of counteraction, and if the bulk of the population is at least neutral. In the decisive first hours or days of revolution a few hundred, or at the most a few thousand, revolutionaries can achieve success even in a large country. (In the eighteenth and nineteenth centuries the pace was usually more leisurely.) Once a strong revolutionary dictatorship has been established, a comparatively small mi-

nority can retain power for a considerable time, provided that its policy does not antagonize all of the people for most of the time or permit the antagonized to organize. Technical progress in transportation and communication has made it much easier for such a regime to influence the whole population and control the entire territory, whereas in the past there was always a danger of counterrevolution breaking out in a remote part of the country.

The timing of a revolutionary attempt is of crucial importance; so are the personalities of its leaders. It is doubtful whether the October Revolution of 1917 would have succeeded without Lenin and Trotsky or indeed whether it would have been undertaken at all. The Bolshevik uprising succeeded, whereas communist or left-wing extremist revolutions elsewhere in Europe (Budapest, Munich) failed. Lenin's party was numerically stronger and better organized than the rather amorphous groups that supported the Bavarian and Hungarian revolutions; the ruling class was weaker and more discredited in Russia than in central Europe. In addition, counterrevolutionary action by foreign powers was, for geographic reasons, less effective in Russia than in Hungary and southern Germany.

A number of revolutions have failed because their leaders lacked the will and ability to persevere in a revolutionary course of action after the first demonstrations or fighting had taken place (Germany in 1848); they feared anarchy and radicalism in their own ranks more than they feared their enemies. Other revolutions failed because they were betrayed before the planned coups were carried out; since preparation is by necessity limited to a fairly small group, the arrest of some leaders can mean a lasting setback for a revolutionary movement.

**Counterrevolution.** Groups of people adversely affected (or likely to be so affected) by a revolution often band together to try to avert it or, if it has already occurred, to overthrow it. Support for counterrevolutionary movements in the nineteenth century usually came from the aristocracy, the clergy, or the higher ranks of the army—all those with a vested interest in preserving the old order. Toward the end of the century, however, and in particular after World War I, certain important changes took place. Since revolutionaries were appealing to the mass of the people and trying to establish, wherever possible, a mass basis of support for their activities, the counterrevolutionaries had to adapt to the new conditions. To have any political influence at all they were compelled to accept a part (sometimes a large part) of the revolutionaries'

program and to adopt their tactics. In Europe only those counterrevolutionary movements that have done so have enjoyed success for any length of time, although overwhelming outside assistance may serve as well. The counterrevolutionary movement usually restricts itself at first to mere demagogic phrasemongering; but gradually it spreads and gains a momentum of its own. The trade unions founded by Colonel Zubatov of the Russian police in 1900 were originally meant to combat the revolutionaries, but in the competition for popular support they gained a measure of independence, and their sponsors lost control over the movement. The fascist and national-socialist movements in Europe in the 1920s and 1930s usually began as counterrevolutionary movements; but the majority of their leaders did not belong to the old order and had no intention of restoring it. In their extreme forms these movements were quasi-revolutionary in character (National Bolshevism in Germany; Mussolini's Republic of Salò, etc.) or nihilistic (Hitler), rather than reactionary. The traditional categories of "revolutionary" and "counterrevolutionary" continued perhaps to make sense when applied to the more rudimentary forms of European fascism, but they became irrelevant in regard to its more advanced forms and to the totalitarian state in general. The Hungarian revolution of 1956 has proved this beyond reasonable doubt.

**Consequences of revolution.** It is characteristic of a revolution (in contrast to a mere *coup d'état*) that it has far-reaching political, social, and sometimes economic or cultural consequences. Revolutions have frequently been the "locomotives of history"; just as frequently they have been quite senseless, causing much unnecessary suffering. Revolutions have liberated peoples and elevated classes; at the same time some of the greatest crimes in history have been committed in the name of revolution. It has been argued that the emancipation of the third estate in France and the transformation of Russia into a democracy were already under way in 1789 and 1917 and that the revolutions merely gave a spectacular acceleration to an inevitable historical process. Since force begets force, however, violent revolution, though intended to overthrow despotism, very often culminates in a new tyranny. Since the idealism and devotion inherent in revolutionary movements thus serve to bring about a new despotism, the movements can hardly be regarded as great successes in terms of what they set out to do. Against this it is argued that the lasting "social achievements" of revolutions are more decisive than any short-lived "political distortion" or terrorism.

The major revolutionary movements of modern times have claimed a world-wide mission and have been expansive in character. The French Revolution in the years 1789-1815 undoubtedly helped spread civil liberties throughout Europe; some of its reverberations were felt in Europe and the rest of the world long after. Yet the attempt to bring the blessings of civil liberties forcibly to other countries produced a chauvinistic, antiliberal reaction in some parts of Europe, which had fateful consequences. Revolutions that set themselves limited objectives may succeed in achieving genuine reforms if they represent the aspirations of the majority of the population. The more violent a revolution, the greater the amount of coercion applied, the more likely that dictatorship, ostensibly established for a short "transition" period only, will be perpetuated. Revolutions that aim at a total transformation of society claim to act in the interests of the majority, but since the majority does not realize where its own best interests lie, it is left to a small avant-garde to make the decisions for it. In consequence it is more than likely that a society will emerge in which severe repression seems to be permanently built in. Such a state may achieve striking results in various fields, such as the national economy or national defense, but to judge from past experience, it will not succeed in building a freer and more just society.

**Doctrines of revolution.** The modern conception of revolution as a fundamental transformation dates back to the eighteenth century, although it is not quite correct to argue that the idea of a radical new beginning was altogether alien to ancient times. However, until the late eighteenth century the advocates of utopian societies were, with some notable exceptions, not revolutionaries, and revolutionaries were not utopians. Ordinarily, the revolutionaries fought against absolute monarchs or tyrants in the name of a natural order that the latter had violated. This tradition was expressed in the American War of Independence and the early stages of the French Revolution. The French Revolution produced an immense amount of theorizing on the subject, and the discussion received fresh fuel from the many revolutionary outbreaks of the nineteenth century. The famous exchange between Burke and Paine set the pattern for an argument which has continued to the present day. The conservative viewpoint rests on the belief that political and social continuity are essential prerequisites for an orderly society; that abstract rights and model constitutions count for little in practice; that revolutions are at best destructive and only replace one despotism by another—if they do not result in a



period of general anarchy. The nineteenth-century conservative regarded the revolutionary movement as inherently sinful, its leaders as misfits and criminals, and its followers as the deluded dregs of society.

Revolutionary theory in its more extreme forms, on the other hand, held that revolutionaries were only the most consistent fighters for freedom, justice, and the other ancient ideals of mankind and that therefore in the long run they could do no wrong. If conservative and liberal doctrine was largely based on a fear of violence, revolutionary theory belittled the significance of violence and terror and almost totally ignored the realities of political power in a modern society.

In recent decades the Marxist-Leninist doctrine of proletarian revolution has all but monopolized discussions on the subject. It claims to be a scientific doctrine and regards history as proceeding according to ascertainable laws. Revolutions fulfill a crucial function in the development of society: they overthrow the old social order, which has outlived its usefulness, and establish a new order; in this process power passes from the hands of one class (which is reactionary) to another (which is progressive in character). According to Marxist theory, the productive forces of society at certain stages of their development clash with the existing relations of production. The existing social order begins to impede their further development, and an epoch of social revolution begins. The basic problem in each revolution is the question of state power; through the class struggle power passes from one class to another. The class struggle may lead to civil war; its very highest form is revolution. All revolutions before 1917 only brought about the replacement of one form of exploitation by another; only the socialist (communist) revolution puts an end to man's exploitation by man. This is the final revolution, for in a socialist society there can be, by definition, only "nonantagonistic conflicts," conflicts that can be resolved in a peaceful way.

The historical fate of Marxist-Leninist doctrine is of great importance. It has been very successful, but not where it expected to succeed—in the industrially developed countries of Europe and North America. Instead it became the ideology of an industrial revolution in backward countries. The revolutionary situation that Marx had expected to develop in western and central Europe failed to arrive. In the underdeveloped countries, on the other hand, the Western impact produced strains that made for a different kind of revolutionary situation. These revolutions were not proletarian in character; neither in China nor in Cuba did the

working class take a prominent part in the revolution. These revolutions, however, brought to power regimes that were either communist from the first (China) or that soon adopted a "Marxist-Leninist" viewpoint (Cuba).

These historical fortunes of the Marxist-Leninist theory of revolution made various ideological mutations and adaptations necessary. The socialist parties of Europe became democratic in character; some of them continued to pay lip service to revolution, but their practice, and gradually their theory, became "revisionist." In the late 1950s several European Communist parties also began to show what some observers regarded as reformist inclinations. In Russia, Trotsky and others had propagated the idea of the permanent revolution, the doctrine that a "bourgeois revolution" in Russia (or a similar country) would immediately be followed by the establishment of a "proletarian dictatorship" and that an upheaval in one country would inevitably lead to revolution on a wider scale. The world revolution failed to materialize, however, and in the Soviet Union there developed what has become known (though never defined in detail) as the Stalinist concept of the "revolution from above." Revolution no longer meant the liberation of society from outworn political and spiritual fetters. Now it stood for the reshaping of society by a dictatorial regime controlling a centralized state apparatus and an all-pervading party organization. The "revolution from above" established communist regimes in eastern Europe; abortive revolutionary uprisings from below (Berlin in 1953, Poznan and Budapest in 1956) were later directed against these regimes.

Revolutionary movements continue to exist in the underdeveloped parts of the world; the revolutionary potential of these countries is far more pronounced than that of the industrially advanced nations. Revolutionary doctrine in these areas has absorbed some elements of the Marxist-Leninist theory, but to a very considerable extent it springs from other sources, such as resentment against the white nations that ruled the colonial world for so long. Attempts to organize the former colonies and underdeveloped countries into one united front against the metropolitan countries can be traced back to the 1920s, and in a somewhat modified form the idea has found new proponents in the 1960s; they argue that the underdeveloped countries are the "proletarian nations," the only truly revolutionary force on a world scale.

According to the Chinese communist thesis, revolutions (and wars of liberation) would and should continue regardless of consequences, whereas most Russian and European communists maintain that

in the nuclear age any large-scale wars of liberation could cause as much destruction as an imperialist war and have therefore become self-defeating as means of achieving political results. They argue that peaceful revolution has become possible as a result of a shift in the global balance of power and the growing attraction of the ideas of communism; the Chinese have violently attacked such views as revisionist.

The historical role of the Marxist theory of revolution has been enormous. With its emphasis on social and economic factors, it helped to put into wider perspective events that previously had been interpreted almost exclusively from a political angle; it has also been more useful than "static" interpretations as a guide to revolutionary situations. However, with the end of the epoch that produced it, this doctrine gradually ceased to be relevant; it ceased to be so in the communist countries because these faced, in the postrevolutionary situation, problems that could not be foreseen in the nineteenth century; it ceased to be relevant in the Western countries because there was virtually no likelihood of a revolutionary situation arising. In the *tiers monde* it continues to be of relevance as a doctrine of industrial revolution—one in which, however, national or local characteristics are of growing importance.

WALTER LAQUEUR

[See also INTERNAL WARFARE; MARXISM; SOCIAL MOVEMENTS.]

#### BIBLIOGRAPHY

- AMANN, PETER H. 1962 Revolution: A Redefinition. *Political Science Quarterly* 77:36–53.
- ARENDT, HANNAH 1963 *On Revolution*. New York: Viking.
- ARON, RAYMOND (1944) 1946 *L'homme contre les tyrans*. 2d ed. Paris: Gallimard.
- ARON, RAYMOND 1951 *Les guerres en chaîne*. Paris: Gallimard.
- BORKENAU, FRANZ 1937 State and Revolution in the Paris Commune, the Russian Revolution, and the Spanish Civil War. *Sociological Review* 29:41–75.
- BRINTON, CLARENCE CRANE (1938) 1952 *The Anatomy of Revolution*. Rev. ed. Englewood Cliffs, N.J.: Prentice-Hall. → A paperback edition was published in 1957 by Vintage.
- BURKE, EDMUND (1790) 1960 *Reflections on the Revolution in France, and on the Proceedings in Certain Societies in London Relative to That Event*. New York: Dutton; London: Dent.
- COHN, NORMAN (1957) 1961 *The Pursuit of the Millennium: Revolutionary Messianism in Medieval and Reformation Europe and Its Bearing on Modern Totalitarian Movements*. 2d ed. New York: Harper.
- DAVIES, JAMES C. 1962 Toward a Theory of Revolution. *American Sociological Review* 27:5–19.
- FRIEDRICH, CARL J. (editor) 1966 *Revolution*. New York: Atherton.

HATTO, ARTHUR T. 1949 Revolution: An Enquiry Into the Usefulness of an Historical Term. *Mind* 58:495–517.

LENIN, VLADIMIR I. (1917) 1964 The State and Revolution: The Marxist Theory of the State and the Tasks of the Proletariat in the Revolution. Volume 25, pages 381–492 in Vladimir I. Lenin, *Collected Works*. 4th ed. Moscow: Foreign Languages Publishing House.

LICHTHEIM, GEORGE 1961 *Marxism: An Historical and Critical Study*. London: Routledge; New York: Praeger.

MANNHEIM, KARL (1939–1943) 1950 *Diagnosis of Our Time*. London: Routledge.

MARX, KARL; and ENGELS, FRIEDRICH (1848) 1959 *The Communist Manifesto*. London: Allen & Unwin. → A paperback edition was published in 1964 by Washington Square Press.

MARX, KARL; and ENGELS, FRIEDRICH (1848–1898) 1962 *Selected Works*. 2 vols. Moscow: Foreign Languages Publishing House.

MÜHLMANN, WILHELM E. 1961 *Chiliasmus und Nationalismus: Studien zur Psychologie, Soziologie und historischen Kasuistik der Umstürzbewegungen*. Berlin: Reimer.

PAINE, THOMAS (1791) 1951 *The Rights of Man*. New York: Dutton.

TOCQUEVILLE, ALEXIS DE (1856) 1955 *The Old Regime and the French Revolution*. Garden City, N.Y.: Doubleday.

TROTSKY, LEON (1930) 1962 *The Permanent Revolution*. London: New Park Publications.

#### REWARDS

See LEARNING, article on REINFORCEMENT; INDUSTRIAL RELATIONS, article on REWARD SYSTEMS AND INCENTIVES; WAGES.

#### RICARDO, DAVID

David Ricardo (1772–1823), English economist, is best remembered for his theory of rent and his theory of comparative cost. He was the first to show that the ultimate incidence of a tax is governed by the laws of economic distribution, laws whose determination he regarded as "the principal problem of Political Economy."

Ricardo was born in London, the third child in a Sephardic Jewish family that had emigrated from Holland in 1760. At the age of 11 he was sent back to Holland to attend a special school, the Talmud Tora, attached to the Portuguese Synagogue in Amsterdam. He returned to London three years later to work for his father, who had become a successful member of the stock exchange. In 1793, against the wishes of his parents, he married the daughter of a Quaker, and they promptly disinherited him. Striking out on his own, he soon accumulated a small fortune as a stockjobber and loan contractor. In 1814, at the age of 42, he retired from business, purchased an estate in Gloucester-



shire, and began to devote himself to literary pursuits.

Ricardo's interest in economics dated back to 1799, when he accidentally came across a copy of Adam Smith's *Wealth of Nations*. Ten years later he made his debut in print with a newspaper article on the bullion controversy, subsequently expanded into a vigorous pamphlet entitled *The High Price of Bullion: A Proof of the Depreciation of Bank Notes* (1810; in *Works*, vol. 3, pp. 47-127). The Bullion Committee report, which appeared the following year, agreed with Ricardo that the current inflation was due to the Bank of England's failure to restrict the issue of currency. The fame of the committee's report lent prestige to Ricardo's tract, so much so that later generations incorrectly credited Ricardo with inspiring the report.

In 1814 he turned his attention to commercial policy and a year later published the important *Essay on the Influence of a Low Price of Corn on the Profits of Stock* (in *Works*, vol. 4, pp. 1-41). It appeared close upon similar essays by Robert Torrens, Edward West, and Thomas Robert Malthus. With varying degrees of clarity, all four pamphlets expressed the idea that was soon to become the cornerstone of classical political economy, the law of diminishing returns to additions of capital and labor applied to land. But only Ricardo went on to develop the implications of this proposition in a major treatise, *On the Principles of Political Economy and Taxation* (1817; see *Works*, vol. 1). The story of how he was urged, goaded, and even bullied into completing this task by his friend James Mill is now well known, owing to Piero Sraffa's recovery of the Ricardo-Mill correspondence, which is included in the *Works and Correspondence* (vols. 6-9).

Further tracts by Ricardo appeared in 1817, 1819, and 1822. In 1821 he published the third edition of the *Principles*, adding a controversial chapter "On Machinery," in which he conceded that technological progress might prove injurious to the working class (see BARTON, JOHN). In 1819 Ricardo obtained a seat in the House of Commons and for the next few years took an active part in parliamentary discussions of current issues. He made a grand tour of the Continent in 1822, keeping a travel journal, which is included in the *Works and Correspondence* (vol. 10, pp. 175-352). He died of a cerebral infection in 1823, at his estate, Gatcomb Park, leaving a wife and seven children. At his death his estate was valued at about £750,000, of which a third was in landed property and the rest in English and French government bonds (the present purchasing power of the estate would be approximately five times as much).

There never was an economist so single-minded in his devotion to "the dismal science" as Ricardo. He did spend some time studying the natural sciences, particularly geology—which then enjoyed something of the glamour that nuclear physics does nowadays—collecting minerals and participating in the management of the Geological Society of London. But with these exceptions, all his intellectual efforts were given to economics, and his correspondence rarely touches on other topics. Politically he was a Benthamite in that he favored universal male suffrage and as much parliamentary reform as was necessary to bring this about. If he had definite philosophical convictions, they were those of British empiricism as embodied in the works of Bacon, Hobbes, and Locke. He became a Unitarian after his marriage but was probably as near to atheism as Bentham was. There is reason to think that Ricardo followed James Mill in advocating the spread of birth control devices among the working class, although, unlike Mill, he never committed himself on the subject in print.

**Economic thought.** The influence of Ricardo's treatise made itself felt almost as soon as it was published, and for over half a century Ricardo dominated economic thinking in Britain. The leading periodicals and even the *Encyclopaedia Britannica* fell into the hands of his disciples; popular literature echoed Ricardian ideas; and Parliament increasingly succumbed to Ricardian policy proposals. Ricardo's appeal rested on his ability to seize hold of a wide range of significant problems with a simple analytical model that involved only a few strategic variables and yielded, after a few elementary manipulations, dramatic conclusions of a distinctly practical nature. In short, he was the first to master the art that later brought success to Keynes.

At the heart of Ricardo's system is the notion that economic growth must sooner or later be arrested owing to a scarcity of natural resources. In the simple version of his model, as expounded in the 1815 essay, the whole economy is a giant farm, distributing its product among landlord, tenant farmer, and hired laborer. The Malthusian tendency of population to increase up to the limits set by the means of subsistence provides a virtually unlimited supply of labor that can be employed at a constant real wage fixed in terms of corn (or wheat). This, of course, is the notorious "iron law of wages." Every worker is equipped with the same amount of fixed capital (say, a spade), which, being combined with every worker in the same proportion, enters as a simple multiple and, hence, can also be measured in terms of corn. In other words, corn is the only output of the giant farm,

and it is also the only input. As the labor force increases, extra corn to feed the extra mouths can be produced only by extending cultivation to less fertile land or by applying additional capital-and-labor to land already under cultivation with diminishing results. The difference between the net corn product per worker on the least fertile land and the constant corn wage per worker goes to the tenant farmer as profit. Since farmers outbid each other for the best land, the real advantages of working superior land accrue to the landowner in the form of rising rents. As more land is taken up, the net produce per worker falls, whereas the real wage remains the same. Obviously, profits per worker decline. At the same time, the corn value of the capital per worker increases because corn is continually becoming more expensive to produce in terms of real resources used up. Divide the falling profits per worker by the rising capital per worker, and it follows that the rate of profit on capital, which supplies the motive for investment, declines. Eventually, capital accumulation must come to an end.

In the *Principles* this simple one-sector model was replaced by a three-sector model, but the argument and the conclusions reached were essentially the same as in the essay of 1815. The three-sector model gave rise to the question how best to measure the variables employed, and the first chapter of the *Principles* turned into a veritable maze for the reader as Ricardo pursued "the invariable measure of value," which seemed forever to elude him. He failed to realize that an index-number problem such as he had posed cannot, in the nature of the case, be perfectly resolved.

The root of the trouble being the declining yield of wheat per acre of land, it is evident that the short-run solution is to import cheap wheat from countries better endowed with fertile land in exchange for low-cost manufactured goods. In the long run other solutions may be looked for: a change in workers' diets to lessen the dependence on wheaten bread; an effort to slow down the rate of growth of population; and, last, an attempt to foster technical change, particularly of the land-saving variety. Ricardo did not neglect these long-term solutions, but, it is fair to say, he dealt with them in a perfunctory manner. The fact is that despite the numerous references to capital accumulation and population growth, Ricardo's model is not really concerned with problems of economic growth in the long run. His interest in practical results—paradoxical in view of the high level of abstraction that he achieved—led him to emphasize the foreign trade solution to the growth dilemma. In so doing he came to attack the exist-

ing corn laws, which protected British wheat farmers by prohibiting foreign wheat except in years of famine prices. Eager to show that Britain would benefit from specializing in manufactured goods and importing the bulk of its food supply, he hit upon the law of comparative cost as a proof of his case.

Adam Smith had taught that it pays a country to concentrate on the production of those goods which it can produce more cheaply in absolute terms than any other country can. Earlier in the eighteenth century a few authors had advanced the wider rule that each country will find it profitable to import those goods which it can obtain for exports at less cost than their domestic production would entail. Almost no one realized that this meant that under free trade all goods are not necessarily produced in countries where their real cost of production is lowest—it might pay a country to import a product even though it could be produced at less cost at home than abroad. The Ricardian doctrine of comparative cost is simply a rigorous restatement of the informal eighteenth-century rule.

Ricardo used a single numerical example, involving two countries and two goods. He was able to demonstrate that it would benefit England to specialize in "cloth," even if Portugal could produce "cloth" and "wine" more cheaply, provided England had a relative cost advantage in cloth. And if Portugal specialized in wine, where it had an even greater cost advantage than in cloth, it turned out that both countries were better off: with given resources, both countries could enjoy more of both cloth and wine because they would reap the advantages of an international division of labor. Here is the fountainhead of all nineteenth-century free trade doctrine! Although the corn laws were not in fact repealed until 1846, Ricardo's writings helped to make free trade a popular objective of British policy. Unwittingly, Ricardo provided the theoretical justification for the long-run solution to the growth problem which Britain actually adopted in the nineteenth century: Britain became the "workshop of the world" and bought most of her food abroad.

Ricardo did much more than state the law of comparative cost. He also saw its implications for international wage and price levels. David Hume had shown earlier that each country's exports and imports are automatically brought into balance under an international gold standard by the effect of gold flows acting on domestic price levels. A country with an export surplus must attract gold from abroad; the gold inflow, by raising all prices within the country, then eliminates the excess of exports over imports. Similarly, an import surplus



leads to a loss of gold and a self-correcting decline in prices. Ricardo now realized that if Portugal has an absolute advantage in both wine and cloth but a great relative advantage in wine, foreign trade with England is possible only if money wage rates in Portugal are higher than those in England. If the wage rate in terms of gold is the same in the two countries, Portugal will not import cloth, since Portuguese consumers can get cloth more cheaply from domestic suppliers. As a result, England will be forced to ship gold to Portugal to pay for wine imports, and this will continue until gold wages and prices in Portugal rise enough to make it profitable for the Portuguese to import English cloth. In general, then, the low-cost country has the higher money wage rate, and hence higher money prices for similar goods. There is, as Ricardo said, "a natural distribution of specie" between the trading nations of the world that tends, in the absence of tariffs, not merely to equilibrate each country's exports and imports, but also to produce such relative price and wage levels between countries as to induce each to produce those goods in which it has a comparative advantage. And although the world is now off the gold standard, it remains true that everyone who has studied Ricardo's argument can understand why a high level of wages in a country does not necessarily prevent that country from competing successfully in international trade.

The same forces that depress the rate of return on capital in Ricardo's model also raise rents, although Ricardo never made it very clear whether he meant rents per acre of land or the rental share of the total product. Be that as it may, Ricardo viewed resources as shifting between agriculture and manufacturing in search of maximum returns, but he never conceived of land itself as shifting among competing uses. When, in the Ricardian model, cultivation is said to be extended, the new land is simply taken up freely, because land is treated not only as if it were completely fixed in supply but also as if it were absolutely specialized in use. It is this conception of land as a free "gift of Nature," having no use other than the production of wheat, that led Ricardo to the conclusion that rent is price-determined, not price-determining; in a famous formula, one of the many with which Ricardo peppered his otherwise pedestrian prose, "Corn is not high because rent is paid, but rent is paid because corn is high." From the notion that rent is not a reward for any economic service rendered by landowners, Ricardo drew the practical implication that the remission of rents from landlords to tenants would not affect the price of wheat or the rate of profit in agriculture. James Mill went a step further and concluded that the

appropriation of rents by the state would not affect agricultural output, a proposition that Henry George made his own when he called for taxing the "unearned increments" in rental values. The popularity of single-tax propaganda in the latter half of the nineteenth century did much to keep the name of Ricardo alive, and via Henry George's *Progress and Poverty* Ricardo even cast a shadow on the formation of the Fabian Society. In a period when the political power of the landowning aristocracy in England still barred the way to effective electoral reform, a theory that attributed economic difficulties to agricultural protection, asserting as a scientific conclusion that "the interest of the landlord is always opposed to that of the consumer and manufacturer," could not fail to take hold of public opinion.

While Richard Cobden and John Bright were adulterating Ricardo's arguments to suit the needs of the great campaign they were waging for immediate repeal of the corn laws, certain radical writers were putting forth labor's exclusive claim to the total national product, a claim which they argued was a simple logical deduction from the labor theory of value. These were the so-called "Ricardian socialists," forerunners of Karl Marx. The label is to some extent a misnomer, because these men, followers of Robert Owen, cared nothing for Ricardo except insofar as he sanctioned the labor theory of value. The notion that profit and rent constitute a deduction from values created by labor alone stems from *The Wealth of Nations*, not from Ricardo's *Principles*. Indeed, if by a labor theory of value we mean a theory that holds that labor is the sole determinant of relative prices, Ricardo cannot be considered a proponent of the theory. It is true that he believed that the ratios in which goods exchange are quantitatively more influenced by relative labor costs than by, say, relative interest charges. It is also true that he held a cost theory of value which left out the influence of demand. But this is still a far cry from what the Ricardian socialists and Marx himself meant by the labor theory of value. It never occurred to Ricardo to deduce anything about the nature of profit from his theory of value, and it is precisely this that separates Ricardo from Marx. But however misleading the common impression that Ricardo was responsible for the genesis of British socialism, the appeal to Ricardo by socialist writers, particularly the tribute that Marx paid him, gave Ricardo a new audience at a time when his influence was in other respects beginning to wane.

**Appraisal.** It is revealing to compare Ricardo to Adam Smith, the founder of classical economics. As a rigorous theorist, Ricardo was obviously far

ahead of Adam Smith. On the other hand, *The Wealth of Nations* contains more in the way of substantial generalizations on the workings of economic systems than does Ricardo's *Principles*—more, perhaps, than any other treatise written in the nineteenth century, with the exception of Alfred Marshall's *magnum opus*. If the problem of economics is the allocation of limited means among unlimited competing ends, as we are sometimes told, then Adam Smith contributed more to economics than did Ricardo; the only place where Ricardo addressed himself specifically to the allocation problem was in the chapter on foreign trade, but here, at any rate, he saw further and deeper than Adam Smith. If the problem of economics is growth and development—a common assertion nowadays—there is again more in Adam Smith than in Ricardo. But if economics is essentially an engine of analysis, a method of thinking rather than a body of concrete results, Ricardo literally invented the technique of economists. His gift for heroic abstractions produced one of the most impressive models, judged by scope and practical import, in the entire history of economic theory.

Ricardo's style of thinking is exemplified by his treatment of taxation, which takes up more than one-third of the *Principles of Political Economy and Taxation*. The discussion is very little connected with the actual taxes prevailing in his time, with the technical problems of tax administration, with the crying need for fiscal reform; instead, it traces the broad economic consequences of taxes in general with respect to the accumulation of capital, the distribution of national income, and the price level. The tax chapters simply set forth the general implications of Ricardo's theoretical scheme, and their value is solely that of demonstrating his superb logical powers.

As long as the corn laws remained in force, the issue of free trade gave significance to Ricardo's doctrines. When repeal came, John Stuart Mill's *Principles of Political Economy*, published two years later, brought new authority to the Ricardian system suitably amended. After 1870, however, most economists turned their backs on what they understood to be the Ricardian theory of value and agreed with William Stanley Jevons that Ricardo had "shunted the car of economic science to a wrong line" (1871). Marx's praise did not enhance Ricardo's reputation among academic economists. But the last decade of the nineteenth century saw a change in attitude, as a number of writers were suddenly struck by the idea that the old-fashioned Ricardian theory of rent was really a special case of a much more general theory. Ricardo had shown that the final dose of labor-and-capital on an in-

tensively used rent-yielding piece of land adds nothing to rent but consists solely of wages and interest, rent being due to the superior productivity of the intramarginal units. Philip Henry Wicksteed, Knut Wicksell, and John Bates Clark realized that there is nothing unique about a no-rent margin; when land is the variable factor and labor-and-capital is the fixed factor, the margin will be a no-wage, no-interest margin. With that insight the general marginal productivity theory of distribution was born, and to Ricardo's other accomplishments must be added that of having anticipated marginal analysis. In recent years concern over the problem of aggregate effective demand has caused many economists to agree with Keynes that "the complete domination of Ricardo's [approach] for a period of a hundred years has been a disaster to the progress of economics" (Keynes [1933] 1956, p. 33). But this is a harsh judgment which supposes that but for Ricardo's magisterial influence, economics would have addressed itself in the past to the macroeconomic problem of unemployment. Ricardo's avowal of the famous "law of markets," asserting a tendency toward full employment equilibrium, was poorly thought out and remained little more than a dogma. As a monetary theorist Ricardo was not even representative of the best work of his own day. Still, the law of comparative cost and the method of comparative static analysis have survived intact. And the central problem that Ricardo posed, namely, how the changes in the relative shares of land, labor, and capital are connected with the rate of capital accumulation, remains the abiding concern of modern economists. In that sense Ricardo is still with us.

MARK BLAUG

[For the historical context of Ricardo's work, see the biographies of HUME; MALTHUS; SMITH, ADAM; TORRENS; WEST; for discussion of the subsequent development of his ideas, see RENT; VALUE, LABOR THEORY OF; and the biographies of CLARK, JOHN BATES; GEORGE; JEVONS; MARX; MILL; WICKSELL; WICKSTEED.]

#### BIBLIOGRAPHY

- ASHTON, THOMAS S.; and SAYERS, R. S. (editors) 1953 *Papers in English Monetary History*. Oxford: Clarendon. → Discusses Ricardian monetary theory.
- BIAUJEAUD, HUGUETTE 1933 *Essai sur la théorie Ricardienne de la valeur*. Paris: Sirey.
- BLAUG, MARK 1958 *Ricardian Economics: A Historical Study*. Yale Studies in Economics, No. 8. New Haven: Yale Univ. Press. → The rise and fall of the school of Ricardo in the first half of the nineteenth century.
- BLAUG, MARK 1962 *Economic Theory in Retrospect*. Homewood, Ill.: Irwin. → Chapter 4 is a treatment of Ricardo's system in modern analytical terms. Also contains bibliographical references.



- FRANKLIN, BURT; and LEGMAN, G. 1949 *David Ricardo and Ricardian Theory: A Bibliographical Checklist*. New York: Franklin. → Contains serious misprints, and the annotation is frequently misleading. It is, however, the only checklist available.
- HOLLANDER, JACOB H. 1910 *David Ricardo: A Centenary Estimate*. Baltimore: Johns Hopkins Press. → Still the best introduction to Ricardo
- JEVONS, WILLIAM S. (1871) 1957 *The Theory of Political Economy*, 5th ed. New York: Kelley Reprints.
- KEYNES, JOHN M. (1933) 1956 Robert Malthus: The First of the Cambridge Economists. Pages 11–38 in John M. Keynes, *Essays and Sketches in Biography*. New York: Meridian. → A paperback edition was published in 1963 by Norton.
- RICARDO, DAVID *Works and Correspondence*. Edited by Piero Sraffa. 10 vols. Cambridge Univ. Press, 1951–1955. → For Ricardo's theory of value, see Sraffa's introduction. Volume 1: *On the Principles of Political Economy and Taxation*. Volume 2: *Notes on Malthus's Principles of Political Economy*. Volume 3: *Pamphlets and Papers, 1809–1811*. Volume 4: *Pamphlets and Papers, 1815–1823*. Volume 5: *Speeches and Evidence*. Volume 6: *Letters, 1810–1815*. Volume 7: *Letters, 1816–1818*. Volume 8: *Letters, 1819–June 1821*. Volume 9: *Letters, July 1821–1823*. Volume 10: *Biographical Miscellany*.
- ST. CLAIR, OSWALD 1957 *A Key to Ricardo*. New York: Kelley Reprints; London: Routledge. → An excellent guide to "what Ricardo really said." Virtually a compilation of Ricardo's opinions.
- SHOUP, CARL S. (1950) 1960 *Ricardo on Taxation: An Analysis of the Chapters on Taxation in David Ricardo's Principles*. New York: Columbia Univ. Press. → Ricardo's views on public finance.
- VINER, JACOB 1937 *Studies in the Theory of International Trade*. New York: Harper. → Discusses Ricardo's contribution.

## RICE, STUART

Stuart Arthur Rice is distinguished for his contributions to the growth of behavioral approaches in social science, to the progress of the United States government's statistical activities, and to bettering the relations between government and business. In particular, he improved statistical reporting, the organization of international statistics, and the collection of statistics in various foreign countries.

Rice received his Ph.D. in 1924 from Columbia University, where he studied primarily under Franklin H. Giddings. Sociological training at Columbia emphasized quantitative research, and Rice produced in the 1920s a series of pioneer studies (e.g., 1924; 1928) which used quantitative methods to test a number of hypotheses regarding political behavior. He hoped his studies would be models for an objective, empirical, value-free science of politics. One hypothesis he investigated

was that political behavior may be primarily the resultant of attitudes. He assumed that attitudes are normally distributed: abnormal distributions result from disturbing factors. He implied that political attitudes lie, for the most part, along one basic continuum of radicalism–conservatism. Along with William F. Ogburn and others, Rice used election statistics to analyze differences between various subgroups, thus anticipating later ecological voting studies. Following the leadership of A. Lawrence Lowell, he also used roll calls to detect differences between particular legislators and legislative blocs.

Another topic which Rice investigated in his early studies was the change over time in such political behavior as voter turnout and party preferences. He pioneered in the use of panel techniques to show the impact of observed stimuli upon the political attitudes of students, as measured by simple questionnaires.

At the time that Rice made his studies of political behavior, empirical research in social psychology was just beginning. L. L. Thurstone had not yet perfected his attitude-measurement techniques. Rice's use of the normal curve as a model of political behavior was soon outdated. More recently, social scientists commanding new data have elaborated his analysis of the areal distribution of political preferences. Again, the sample survey of the general population has replaced Rice's indirect inferences with the direct measurement of both actual voting and the individual characteristics of citizens. Yet, considering the tools available in the 1920s, Rice made a notable attempt to show the possibilities of an empirical approach to the study of political behavior.

Rice compiled, under the auspices of the Social Science Research Council, a critical review of the methods employed in the social sciences (1931). Case studies of the methods employed in outstanding contributions to social science were arranged in a theoretical setting designed to portray different methodological approaches, without regard to traditional boundaries in subject matter. The basic classification of methods was devised by Rice, as chief investigator and editor, with the assistance of Harold D. Lasswell.

Rice then turned from research and teaching to administrative work in the federal government. Here he did much to professionalize the activities of the Bureau of the Census. He also made notable contributions to the development of social statistics in the United States, the study of the effects of unemployment, the analysis of world standards of living, the establishment of a rationalized sys-

tem of federal statistics, and the projecting of the statistical needs of the United Nations.

Rice was born in northern Minnesota in 1889. He attended the University of Washington, obtaining his B.A. there in 1912 and his M.A. in 1915, and was employed for some years in welfare administration at the local, state, and regional level. After an unsuccessful venture into minor party politics in 1920, he became interested in exploring the rational character of social movements and pursued the objective analysis of sociological data at Columbia University. He taught at Dartmouth College from 1923 to 1926 and at the Wharton School of Finance and Commerce of the University of Pennsylvania from 1926 to 1933.

His first appointment in Washington was as assistant director of the Bureau of the Census, a position he held until 1936. As president of the American Statistical Association in 1933, he was influential in creating the advisory services which in turn produced the central administrative structure for the development and coordination of the federal statistical system. For nearly two decades he was director of the Office of Statistical Standards of the Bureau of the Budget, charged by law with control of statistical work within the government. He won wide praise from the business community for his successful endeavors to reduce the reporting burdens imposed by federal questionnaires. At the beginning of World War II, Rice instigated the creation of the Inter-American Statistical Institute, which supplies the statistical staff of the Pan American Union. At the end of the war, he was active in the establishment of the Statistical Office of the United Nations and served as the first chairman of the UN Statistical Commission, where he displayed skill and ingenuity in working toward greater adequacy and comparability of world statistics.

As president of the International Statistical Institute from 1947 to 1953, he promoted collaboration among the world's statisticians and at the same time pointed out that the conceptions of statistics in the Soviet Union were an impediment to the growth of international statistics. As head of the Statistical Mission to Japan, he helped develop the statistical services of that country. Upon retiring from federal service in 1954, Rice organized and became president of the Surveys & Research Corporation, a consulting firm offering economic, statistical, and management services to United States and foreign governmental agencies, as well as to private industry and nonprofit organizations.

HAROLD F. GOSNELL

[For the historical context of Rice's work, see the biographies of GIDDINGS; OGBURN; LOWELL; for discussion of the subsequent development of Rice's ideas, see GOVERNMENT STATISTICS; POLITICAL BEHAVIOR.]

#### WORKS BY RICE

- 1923 *The Effect of Unemployment Upon the Worker and His Family*. Pages 99-109 in *Business Cycles and Unemployment: Report and Recommendations of a Committee of the President's Conference on Unemployment*. New York: McGraw-Hill.
- 1924 *Farmers and Workers in American Politics*. Columbia University Studies in History, Economics and Public Law, Vol. 113, No. 2. New York: Columbia Univ. Press.
- 1928 *Quantitative Methods in Politics*. New York: Knopf.
- 1930 AMERICAN STATISTICAL ASSOCIATION, COMMITTEE ON SOCIAL STATISTICS *Statistics in Social Studies*. Edited by Stuart A. Rice. Philadelphia: Univ. of Pennsylvania Press.
- 1931 SOCIAL SCIENCE RESEARCH COUNCIL, COMMITTEE ON SCIENTIFIC METHOD IN THE SOCIAL SCIENCES *Methods in Social Science: A Case Book . . .* Edited by Stuart A. Rice. Univ. of Chicago Press.
- 1932 *The Field of the Social Sciences*. Pages 613-632 in James H. S. Bossard (editor), *Man and His World*. New York: Harper. → Bossard's book includes three other articles by Stuart Rice, on pages 633-669.
- 1933 RICE, STUART et al. *Next Steps in the Development of Social Statistics*. Ann Arbor, Mich.: Edwards.
- 1933 WILLEY, MALCOLM M.; and RICE, STUART *Communication Agencies and Social Life*. New York: McGraw-Hill.

#### RICHARDSON, LEWIS FRY

Lewis Fry Richardson (1881-1953), British meteorologist and student of the causes of war, was born at Newcastle-on-Tyne, the youngest of seven children in a Quaker family. He attended the Newcastle Preparatory School, then Bootham School at York. His inclination toward science seems to have been inspired by J. Edmund Clark, a master at Bootham School, who was a meteorologist. However, while still in his teens, Richardson was convinced that "science ought to be subordinate to morals." After leaving Bootham in 1898, he attended Durham College of Science at Newcastle, then King's College at Cambridge.

Several appointments followed, each of short duration. Richardson worked as an assistant in the National Physics Laboratory, as chemist for a peat company, and as director of the physical and chemical laboratory of a lamp company. At the outbreak of World War I he was superintendent of Eskdalemuir Observatory of the Meteorology Office. Thenceforth, meteorology became one of his abiding scientific concerns, and he contributed some thirty papers and a book (1922) to that field.



In 1926 his scientific achievements were recognized by his election as fellow of the Royal Society.

For Richardson, Quakerism is firmly identified with pacifism. Accordingly he declared himself a conscientious objector, a stand that subsequently barred him from university appointments. Nevertheless he participated directly in the war in a non-combatant capacity, as a member of the Friends' Ambulance Unit, attached to the 16th French Infantry Division. After the armistice he returned to the Meteorology Office, but in 1920, when that office became part of the Air Ministry, Richardson resigned his position. He next took charge of the physics department at Westminster Training College and in 1929 became principal of Paisley Technical College and School of Art, his last post. Richardson retired in 1940 in order to devote all of his time to the study of war, searching both for its causes and for means to prevent it [see WAR, *article on THE STUDY OF WAR*].

It is with his work on war that Richardson's name is most frequently associated. Although his pacifist convictions were surely a source of his dedication to this study, on which he spent at least 35 years, it is likely that his deep involvement with meteorology influenced his method of research. The prediction of weather is notoriously difficult, even though the determinants of weather are entirely understood. The principles governing the several variables—the motion of air masses, the changes of pressures and temperatures, the onset of precipitation—are all known, but the interactions among all these factors are so complex that even if the atmospheric conditions over the globe were precisely known at a given moment, the calculation of future states for even a short period would be a superhuman task. At one time Richardson estimated (1922) that it would take 60,000 human computers working at high speed to compute tomorrow's world weather charts before tomorrow's weather arrived. The vital lesson that Richardson therefore learned from the problems of meteorological prediction—one that has been confirmed by modern computing technology—is that events which seem to be governed by chance (as does the weather to one ignorant of the dynamics of air masses) are in fact governed by laws and can be predicted if enough information can be processed.

The link between this insight and Richardson's approach to the phenomenon of war is in his rejection of the idea that war is a rational, or at least a purposeful, form of behavior, as is often assumed in the conventional political conception of international relations. Richardson viewed war

instead in Tolstoyan fashion, as a massive phenomenon governed by forces akin to the forces of nature, over which individuals have little or no control. Accordingly, he ignored all those intricacies of diplomatic-strategic analysis usually pursued by political historians and turned his attention to quasi-mechanical and quantifiable processes which, he assumed, govern the dynamics of the international system of sovereign states. Neither the contents of memoranda and ultimatums nor dynastic claims, territorial ambitions, and networks of alliances play an explicit role in Richardson's theory of war. Instead, one finds differential equations purporting to represent the interactions among states or the spread of attitudes and moods (like the spread of communicable diseases) among the populations of those states. The equations are quite similar to those representing physical or chemical systems, at times tending toward equilibria, at times moving away from equilibria at accelerated rates and culminating in explosions.

Richardson harbored no illusions concerning the adequacy of his mathematical theory in providing an "explanation" of war, much less a means of preventing it. In omitting strategic calculations, considerations of prudence, and other "rational" factors as determinants of war or peace, he was not asserting the irrelevance of these factors. Rather, he sought to build a viable theory of war *in vacuo*, as it were, an admittedly crude but tractable model upon which a more sophisticated theory could be developed. "The equations are merely a description of what people would do if they did not stop to think," he wrote (1960a, p. 12). Some of his equations did fit what actually happened in the years preceding World War I, and Richardson concluded that the Great Powers in that instance were acting as if they were driven by mechanical forces.

In spite of the formal, mechanistic character of the equations that Richardson proposed as a model of international relations, he thought of the causes of war as primarily psychological. The underlying psychology was that of a mass, much simplified by the averaging out of the many opposing pressures, devoid of self-insight or foresight; it was not the psychology of an individual, with a large range of choices, moral convictions, and idiosyncratic preferences. Richardson's first paper in this vein was entitled "Mathematical Psychology of War." Written in 1919 and privately printed, it was not published until 1935. In the intervening years Richardson studied psychology as an external student of University College in London, receiving the special B.Sc. degree in 1929. His work in psychology is

strongly oriented toward the quantitative approach, and some of it shows the same influence of the "meteorological orientation" as do his studies on war (1937).

Richardson published the complete version of his mathematical theory of war, *Generalized Foreign Politics* (1939), on the very eve of World War II. Its point of departure is a pair of differential equations representing a hypothetical interaction between two rival states. The components of interaction are (1) mutual stimulation of armaments buildup, assuming that each nation's rate of increase of armaments expenditures is positively proportional to the other nation's current expenditures; (2) self-inhibition of armaments, assuming that the rate of change of armaments expenditures is negatively proportional to the already existing armaments burden; and (3) constant stimulants to armaments buildup in the form of grievances or ambitions of states. These latter may also be negative, in which case they are interpreted as reservoirs of good will. Cooperation between the rival states, for example, in the form of trade, is interpreted as negative armaments expenditures.

Richardson then examined the dynamics of the postulated system. The solution of the equations (given the initial conditions and the parameters of interaction) determines the time course of the armaments expenditures. By substituting selected values for the parameters, Richardson was able to obtain a good fit of the predicted time course for the armaments buildup of the rival European blocs (the Central Powers and the Entente) in the years preceding World War I. However, the large number of free parameters and the small number of points representing the time course make this "corroboration" of the theory less than impressive.

Of considerably greater interest is the theoretical result deduced from the model, namely that depending on the relative magnitudes of the stimulation and inhibition parameters (but not of the grievance-good-will parameters), the system may be either stable or unstable. If the parameters are such that the system is stable, then an armaments balance is possible (this might also be interpreted as arms control). However, if the parameters are such that the system is unstable, then such a balance is not possible. The system must move one way or the other, depending on the initial conditions, either toward total disarmament and beyond to ever-increasing cooperation or into a runaway race, presumably followed by war.

By noting the rate of disarmament of Great Britain following World War I and the rate of rearmament of Germany prior to World War II,

Richardson was able to get rough estimates of the parameters in question. He concluded that the parameters were well within the region of instability and, moreover, that the initial conditions prior to World War I made it touch and go whether the system would move toward peace or war. Possibly just a slightly lower armaments level or just a little more interbloc trade would have pushed the system toward a united Europe instead of toward world war [see DISARMAMENT].

Following the analysis of the arms race of 1908-1914, Richardson attempted to analyze the similar process that started shortly after Hitler's rise to power. The disappearance of the gold standard as a basis for the measure of expenditures and the scantiness of data from the U.S.S.R. made the analysis of the arms race preceding World War II extremely difficult, beyond anything that a lone investigator could accomplish in a lifetime. Still, Richardson's theory of the mutually stimulating arms race did point to a second world war.

Whether or not these conclusions ought to be taken seriously is a difficult question. Certainly, controlled experiments on the scale of international relations cannot be brought to bear on the critical aspects of Richardson's theory. Yet whatever the explanatory or predictive merits of the theory, one cannot deny that it invites us to see the phenomenon of war from an unusual point of view. This point of view may have been stated earlier (Richardson cites Thucydides as a proponent of the mutual stimulation theory of arms races and wars), but the quantitative implications of rigorously formulated models based on this view seem never to have been worked out.

Besides the two-nation problem of mutual stimulation in an arms race, Richardson also posed the N-nation problem. Again, the relevance of his results to real international dynamics is an open question because of the vastly simplified assumptions on which his models are based. Nevertheless, the results are interesting, not because of the answers they provide but because of the questions they raise. Thus, Richardson found, for example, that in an arms race involving three nations, the situation can be stable for each of the three pairs separately but unstable for all three taken together. This result may be relevant to the currently acute N-nation nuclear force problem [see NUCLEAR WAR].

Following his retirement in 1940, Richardson started extensive empirical investigations. He performed the monumental labor of gathering a vast variety of data related to all "deadly quarrels" that have been known to occur since the end of the



Napoleonic Wars. Richardson's treatment of the data again reveals his view of war and of violence in general as something with which the whole human race is afflicted. This view is diametrically opposed to that represented by strategic thinkers and most clearly expressed by Karl von Clausewitz, who saw war as an instrument of national policy and a normal form of intercourse among civilized states [see CLAUSEWITZ].

To Richardson, war is a particular case of a "deadly quarrel," defined as a violent encounter among human beings resulting in one or more deaths. He placed all such encounters on a scale of magnitude (defined by the logarithm of the number of dead). Thus, single murders appear on this scale as deadly quarrels of magnitude 0 ( $\log_{10} 1 = 0$ ), small riots with some ten victims as deadly quarrels of magnitude 1 ( $\log_{10} 10 = 1$ ), and so on. The two world wars appear on this scale as deadly quarrels of magnitude 7.

Richardson sought to establish a relation between the magnitudes of deadly quarrels and the relative frequency of their occurrence, analogous to the relations established by George K. Zipf (1949) between ranks and sizes of a great variety of objects [see RANK-SIZE RELATIONS]. But unlike Zipf, who singled out the rank-size relation as a unifying principle of fundamental importance, Richardson treated this relation as one of many to be examined in the search for regularities from which, he hoped, the laws governing human violence would emerge.

Richardson sought to relate the frequencies of wars not only to their magnitudes but to every other conceivable factor that could be extracted from the data. He examined the effect of the existence of common frontiers between the combatants and of the existence of a common language, of a common religion, and of a common government (as in civil wars). He made a list of all the "pacifiers" that have been supposed at one time or another to counteract the tendency to violent outbreaks, such as distraction by sports, diversion of hatred to other groups, the direction of hatred inward, armed strength as a deterrent, collective security, or intermarriage among potentially hostile groups. None of the indices related to these pacifiers shows up as a significant contributing factor either to the likelihood of a "deadly quarrel" or to its prevention. Possible exceptions are international trade and allegiance to a common government. Thus, hardly anything in the way of new knowledge as to the "causes" of wars has emerged from this monumental analysis, unless one views as new the refutation of established notions by negative

results. In particular, neither armed might nor collective security measures (contrary to widespread opinion) emerge as significant war-preventing influences.

The failure of the "statistics of deadly quarrels" to generate a theory on the causes of wars was to be expected in view of the magnitude of the problem. Because Richardson worked alone and had no access to modern computing machinery, the bulk of his effort was absorbed in tedious data gathering and routine calculations. It is likely, of course, that any crudely empirical brute-force attack on the causes of wars is inherently doomed to failure. The nature of the primary contributing causes may be shifting rapidly and may be quite different in different cultural milieus, so that lumping together all the deadly quarrels in the world for a period of some 120 years may be statistically meaningless. Still, the pioneering significance of this work and Richardson's gallantry in attacking the formidable problem singlehanded should not be underestimated. Whatever conclusions he drew or failed to draw, his work remains a rich collection of data.

Two of Richardson's books were published in 1960, under the titles of *Arms and Insecurity* and *Statistics of Deadly Quarrels* (1960a; 1960b). The former, edited by Nicolas Rashevsky and Ernesto Trucco, contains Richardson's mathematical theories of arms races, worked out in great detail, and the data relevant to those theories. The latter, edited by Quincy Wright and C. C. Lienau, contains the analysis of the voluminous data on violence ranging from murders to world wars. The two books were widely reviewed and helped establish Richardson's reputation as the pioneer of research into the causes of war.

Following the publication of these two books, a number of investigators in England and the United States undertook to combine Richardson's methods of mathematical model construction and data analysis in the field of international relations. In particular, content analysis methods have been used in attempts to obtain indices of international hostility; various attempts have been made to extend Richardson's theory of arms races to cover the present period; and his methods of correlation analysis, designed as a search for the causative factors of wars, have been modified and refined. In Volume 8 of *General Systems* some of these investigations were published or reprinted under the heading "After Richardson" (Rummel 1963; Smoker 1963).

The principal value of Richardson's contribution lies in the idea of bringing quantitative analysis to

bear upon the possible massive system-dynamic determinants of war. His work stands at one end of a spectrum, the other end of which is represented by game-theoretic analysis, where it is supposed that international relations can be viewed as interplays of strategies calculated by "rational players." This latter view is by far the more prevalent among political scientists. Richardson's quasi-deterministic view of international relations is complementary to the strategic view, which assumes rationality in the pursuit of "interests" but leaves unanalyzed the genesis of the interests. The strategic view may inquire how nations conduct (or would conduct, if they were rational) a diplomatic-military game but says nothing about how the game got started, why enmities are built up between some states and not between others, or, of course, why states behave so frequently and so clearly against their own interests. Although Richardson did not shed much direct light on these matters, his approach raises important questions that are too often ignored in the purely diplomatic-military approach to international relations.

#### ANATOL RAPOPORT

[See also DISARMAMENT; PEACE; WAR. Other relevant material may be found under INTERNATIONAL RELATIONS; SYSTEMS ANALYSIS.]

#### WORKS BY RICHARDSON

- 1922 *Weather Prediction by Numerical Process*. Cambridge Univ. Press.
- 1933 The Measurability of Sensations of Hue, Brightness or Saturation. Pages 112-114 in *Physical Society, London, Report of a Joint Discussion on Vision*. London: The Society.
- 1935 Mathematical Psychology of War. *Nature* 135: 830-831; 136:1025 only.
- 1937 Hints From Physics and Meteorology as to Mental Periodicities. *British Journal of Psychology* 28:213-215.
- 1939 *Generalized Foreign Politics: A Study in Group Psychology*. Cambridge Univ. Press.
- 1960a *Arms and Insecurity: A Mathematical Study of the Causes and Origins of War*. Edited by Nicolas Rashevsky and Ernesto Trucco. Pittsburgh: Boxwood; Chicago: Quadrangle.
- 1960b *Statistics of Deadly Quarrels*. Edited by Quincy Wright and C. C. Lienau. Pittsburgh: Boxwood; Chicago: Quadrangle.

#### WORKS ABOUT RICHARDSON

- BOYCE, ANNE O. 1889 *Records of a Quaker Family: The Richardsons of Cleveland*. London: Harris.
- GOLD, E. 1954 Lewis Fry Richardson 1881-1953. Royal Society of London, *Obituary Notices of Fellows* 9:216-235. → Includes 4 pages of bibliography.
- RAPOPORT, ANATOL 1957 Lewis F. Richardson's Mathematical Theory of War. *Journal of Conflict Resolution* 1:249-299.
- SHEPPARD, P. A. 1953 Dr. L. F. Richardson, F.R.S. *Nature* 172:1127-1128.

#### SUPPLEMENTARY BIBLIOGRAPHY

- HORVATH, WILLIAM J.; and FOSTER, CAXTON C. 1963 Stochastic Models of War Alliances. *General Systems* 8:77-81.
- RUMMEL, RUDOLPH J. 1963 Dimensions of Conflict Behavior Within and Between Nations. *General Systems* 8:1-50.
- SMOKER, PAUL 1963 A Mathematical Study of the Present Arms Race. *General Systems* 8:51-59.
- ZIFF, GEORGE K. 1949 *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology*. Reading, Mass.: Addison-Wesley.

#### RISK

See DECISION MAKING, article on ECONOMIC ASPECTS; ECONOMIC EXPECTATIONS; PROFIT.

#### rites of passage

See MYTH AND SYMBOL; RITUAL; and the biography of GENNEP.

#### RITTER, CARL

In his own time Carl Ritter (1779-1859) was one of the best-known German scholars in the world. For the few who are still aware of him, his voluminous writings of more than 25,000 pages are a mine of facts and perspectives. The most widely read biography, written by his brother-in-law, Gustav Kramer (1864-1870), considered him in strictly theological terms; and as a consequence, the Positivists who followed condemned him, without examining whether his religious beliefs in fact distorted his scientific work. Nevertheless, Ritter, with Alexander von Humboldt, is still generally considered one of the founders of modern scientific geography.

Ritter's life encompassed the German Enlightenment, romanticism, and the Biedermeier era, and—more clearly even than in Humboldt's case—his work mirrored his active participation in all the intellectual movements of his time. He considered it providential that he found himself repeatedly in the center of contemporary intellectual controversies.

His father, an outstanding diagnostician who was personal physician to the sister of Frederick the Great, died in 1784, leaving his widow, a devout Pietist, with six children and in straitened circumstances. In the same year, the well-known pedagogue Christian G. Salzmann had founded his Philantropinum in Schnepfenthal; since he was unable immediately to find pupils, he admitted on scholarships two boys who seemed to be highly talented—Carl Ritter and his brother Johannes, as well as their tutor, Guths Muths. For 11 years,



longer than any other pupil, Carl remained at that institution of sober enlightenment, which neglected humanistic studies—stressing instead the development of such tangible values as a sound body and a good character—and which was designed to produce not scholars but citizens.

Since Ritter showed a talent for drawing, Salzmann planned to make him a copperplate engraver, but chance gave his life a different direction. The Frankfurt banker Bethmann, an ancestor of the German Chancellor Bethmann-Hollweg, happened to be in Schnepfenthal. He hired Ritter as a tutor for his two boys, first sending him for a few years to the University of Halle to fill in any gaps in his education in general and to study cameralism in particular. The Bethmanns remained Ritter's benefactors until he became self-supporting.

Ritter's main initial interest was pedagogy, which brought him into close contact with Johann Heinrich Pestalozzi. He became not only one of the most knowledgeable interpreters and critics of Pestalozzi's "method" but was also chosen by Pestalozzi to succeed him at Iferten; when Ritter refused, his friendship with Pestalozzi came to an end.

Even in his pedagogic period, which lasted from his entry into the Bethmann household until the start of his studies in Göttingen (from 1798 to 1813), Ritter had become interested in geography. His early work, *Europa: Ein geographisch-historisch-statistisches Gemälde* (1804–1807), was still completely under the influence of the then-current statistical political science. However, in the atlas *Sechs Karten von Europa* (1804–1806) that accompanied his work, he revealed his ability to envision the continent as a whole. These maps conceive of the elements of an area as layers: the surface, vegetation, cultivated plants, animal life, population and tribes, etc. He intended the atlas merely as an educational aid for schools, yet it was the origin of the kind of cartographic documentation that is commonly used today.

Ritter's thinking and his scholarly aims were strongly influenced by travels through the magnificent Swiss landscape and by his contacts with Pestalozzi and the theologian Johannes Niederer. He came not only to appreciate the individual character of various kinds of scenery and to develop the idea of interaction between nature and culture, but also to see it as his mission to explore this idea of interaction for the benefit of mankind.

Pestalozzi fostered these general ideas by acquainting Ritter with the idea of a "type," and by suggesting that this concept of type be introduced into the field of geography in a textbook for his

school. There was a strange contradiction in Pestalozzi's predilection for types and his radical condemnation of definitions, which he considered as abstract and therefore lifeless and rigid. For Pestalozzi the only adequate definition was a complete description, and Ritter's later works show that he accepted this view.

Niederer awoke Ritter from his cool, rationalistic moralism to a deeply felt Christianity, though he did not at once become a member of a particular church. Ritter searched for a suitable intellectual foundation for his Christian views, for instance in the works of Friedrich Schleiermacher, and only much later, in Berlin, did he become a Lutheran.

When Ritter submitted to the famous geologist Leopold von Buch the manuscript of the textbook on geography that he had promised to Pestalozzi, it was criticized so severely in its details that Ritter scrapped it. Instead he undertook to incorporate all the known details of geography in a vast description of the entire earth, hoping ultimately to produce the general geographic textbook suggested by Pestalozzi.

In 1813 Ritter went to the University of Göttingen as a tutor to the Bethmann boys and to pursue his own academic plans. Göttingen was at the time the only intellectually free institution in Germany, where research without censorship or tutelage was possible, and it was therefore the gathering place for an elite. Although Ritter did not participate actively in university affairs, he came into contact with the leading men there, such as the anatomist and polyhistorian Johann Friedrich Blumenbach, who played a role also in the development of geography and anthropology; the historian Arnold Heeren; the anatomist Samuel Sömmering; and the mineralogist Hausmann. His contacts extended to Berlin, where the jurist Savigny and the theologian Schleiermacher were his friends. All of these, as well as earlier friends, such as August von Schlegel, Baron Karl vom und zum Stein, and Madame de Staël, were connected with the romantic movement.

In response to new intellectual currents, Ritter studied the entire literature on geography from ancient times to contemporary travel descriptions and made excerpts from it, in order to gain for geography a clearly defined sphere within the sciences and a body of problematics. The principal subject matter for further study was, as he saw it, the surface of the earth from the time that man began both to influence the development of the world and to be influenced by it. This means that geography for Ritter was the science of the surface of the earth in its material utility, and especially

in its interaction with the people who populate it and develop their cultures on it.

The year 1817 saw the publication of the first volume of his *Die Erdkunde im Verhältniss zur Natur und zur Geschichte des Menschen* . . . ; after he had completed the second volume (1818), he began a second edition, which by the time of his death had grown to 21 large volumes dealing with all of Africa and Asia, except for a small western portion of Asia Minor.

The originality of this book soon was acclaimed all over the world, and it also attracted the attention of the Prussian general staff and of those who were responsible for the re-establishment of the University of Berlin. Both groups joined in calling Ritter to Berlin in 1820, to teach geography at the university and to instruct the Prussian military leadership corps. His success in teaching was considerable, although its full impact cannot be judged without further research. The most important German generals, especially Roon and Moltke, were his closest students; they wrote valuable textbooks and reports and applied in practical strategy what Ritter had taught them. At the university, he found enthusiastic and productive disciples among many historians and among such geographers as Heinrich Kiepert, J. E. Wappäus, Heinrich Berghaus, Heinrich Barth, and Elisée Reclus. For all his self-confidence Ritter was modest and reserved, and his public appearances were confined to numerous congresses at home and abroad; he disdained taking advantage of his position as instructor in geography and history to the Prussian princes and princesses to further his personal social position. Nevertheless, the intervention of friends, particularly that of Alexander von Humboldt with the king, permitted him almost every year to undertake long and at that time still very difficult travels throughout nearly all of Europe. Beautifully written travel reports (see G. Kramer 1864–1870, part 2) give testimony to his skill in observing scenery and people and disprove the general opinion that he was a mere bookworm. He was active until he died at the age of 80, soon after his friend Humboldt.

**Evaluation.** What remains of Ritter's work is, first of all, his capacity to perceive a geographic entity and geographic relationships and so to develop the study of regions instead of the traditional study of states. The fluctuating boundaries between states meant little to him, since he described each area in terms of the actual changes that have gradually been established by science. He sought to develop a new and more meaningful organization of the earth into regions, based on intrinsic sources of coherency. These regions might be, for example,

river basins, plateaus, or terraced areas. They are parts of the larger units of the total earth organism, that is, of the continents. To establish these lesser units, it is necessary to have an accurate topographical map and as complete a description as possible of the landscape in a given area, as well as of the climate, the vegetation, the animal life, and finally of man and his historical interaction with the foregoing elements. Clearly the fundamental idea is that all these elements are interconnected and that their differential regional occurrence and interaction determines the character of a particular area.

The human population of a specific area was what aroused Ritter's particular interest. He considered man the fulfillment of creation and its ultimate purpose. Geography, he believed, would fail in its central task if it did not understand how man influences the space in which he lives and is in turn "educated" by the advantages and obstacles that the space offers. Ritter devoted special attention to the period of the greatest cultural development of each area, since he believed that in such periods the highest degree of harmony between nature and culture can be assumed to exist. Thus he became the founder of the kind of historical study of regions that only recently has been seriously revived.

Geographers no longer adhere to Ritter's method of revealing to the reader all the sources of their material. Ritter has been criticized for succumbing to the weight of his sources, for not being able to discriminate the essential. This criticism overlooks the fact that it was he who introduced into geography the critical review of contradictory sources and that he presented clearly the concept of the persistence or change of certain elements in the geographic-historical development of the areas he described.

The romantic principle of "polarity," long since abandoned by geography, permitted Ritter to discover such polar relationships as those between the hemispheres of land and water, between central and peripheral areas, between oceanic and continental phenomena in all their possible natural and historical variations, and between the climatic zones within the polaric tension of north and south. Also, from romanticism came Ritter's concepts of "individuality" as applied to units of space and of "organism," applied to the earth as organism. In Ritter's view this meant more, however, than that the whole is greater than the sum of its parts—an idea that has been revived by the new science of geography in its concern for the "integration" of all parts.



Ritter took a unique position with regard to the natural philosophical conception of the earth as based on a plan that was conceived at creation and that classifies all things according to their endowed ability to develop. (This position is fully developed by Henrick Steffens in his *Anthropology*, published in 1822.) Ritter did not reject natural philosophy a priori as did Humboldt and the majority of scholars at that time; indeed, he was rather inclined toward it. But he refused to acknowledge its claim to being a superior kind of science, based on the specialized sciences it scorned. Avoiding both extremes, he endeavored to test natural philosophy, thereby building a bridge between it and factual research. His scientific integrity kept him always nearer to the facts than to lofty "speculations." He likened himself to a "thinking anatomist or physiologist" who investigates the structure of the earth organism, critically recording and evaluating all facts in the hope of discovering the underlying plan ("no matter whether it be divine or natural"). He did not conceal facts, nor did he abuse them; but he did expect that after a mass of detail was digested he would find a system that would not only permit an interpretation of the earth in mechanical terms but would also be meaningful.

Ritter was much more critical and cautious than many of today's scholars, who are inclined to generalize and to establish rules and laws. His belief in a world order (divine or natural) never led him to a distortion or misrepresentation of facts, and it served him as an incentive for finding innumerable connections which before him the "meaningless description of the earth could not have imagined." Thus he sought to justify himself vis-à-vis the kind of science that developed around him, which considered it its first duty to divest the world of myth; and he provoked the accusation that he was a teleologist, which has dimmed his reputation as a scientist to the present day.

ERNST PLEWE

[See also REGION.]

#### WORKS BY RITTER

- (1804-1806) 1820 *Sechs Karten von Europa*. Schnepfenthal (Germany): Erziehungsanstalt.  
 1804-1807 *Europa: Ein geographisch-historisch-statistisches Gemälde*. 2 vols. Frankfurt am Main (Germany): Hermann.  
 (1817-1818) 1822-1859 *Die Erdkunde im Verhältniss zur Natur und zur Geschichte des Menschen: Oder allgemeine, vergleichende Geographie*. 21 vols. and 2 index vols. 2d ed. Berlin: Reimer.  
 1820 *Die Vorhalle europäischer Völkergeschichten vor Herodotus, um den Kaukasus und an den Gestaden des Pontus: Eine Abhandlung zur Alterthumskunde*. Berlin: Reimer.

- 1831 RITTER, CARL; and O'ETZEL, F. A. *Hand-Atlas von Afrika*. Berlin: Reimer. → Contains 14 maps.  
 (1852) 1863 *Geographical Studies*. Boston: Gould & Lincoln. → First published as *Einleitung zur allgemeinen vergleichenden Geographie*.  
 1861 *Geschichte der Erdkunde und der Entdeckungen*. Vorlesungen an der Universität zu Berlin. Berlin: Reimer.  
 (1862) 1881 *Comparative Geography*. New York: Bragg. → First published as *Allgemeine Erdkunde*.

#### SUPPLEMENTARY BIBLIOGRAPHY

- BITTERLING, R. 1929 Carl Ritter zum Gedächtniss. *Geographischer Anzeiger* 30:233-264.  
 Carl Ritter zum Gedächtniss. 1959 *Die Erde: Zeitschrift der Gesellschaft für Erdkunde zu Berlin* 90:97-254.  
 GAGE, WILLIAM L. 1867 *The Life of Carl Ritter, Late Professor of Geography in the University of Berlin*. New York: Scribner.  
 GUYOT, ARNOLD H. 1860 *Carl Ritter: An Address to the American Geographical and Statistical Society*. Princeton, N.J.: Privately printed.  
 HÖZEL, EMIL 1896 *Das geographische Individuum bei Karl Ritter und seine Bedeutung für den Begriff des Naturgebietes und der Naturgrenze*. Leipzig: Teubner.  
 KRAMER, FRITZ L. 1959 A Note on Carl Ritter: 1779-1859. *Geographical Review* 49:406-409.  
 KRAMER, GUSTAV (1864-1870) 1875 *Carl Ritter: Ein Lebensbild nach seinem handschriftlichen Nachlass dargestellt*. 2d rev. & enl. ed. 2 vols. Halle (Germany): Verlag der Buchhandlung des Waisenhauses.  
 PFEIFER, GOTTFRIED 1960 Ritter, Humboldt und die moderne Geographie. Pages 69-83 in *Deutscher Geographentag*, 32d, Berlin, *Tagungsbericht und wissenschaftliche Abhandlungen*. Wiesbaden (Germany): Steiner.  
 PLEWE, ERNST 1960 Carl Ritters Stellung in der Geographie. Pages 59-68 in *Deutscher Geographentag*, 32d, Berlin, *Tagungsbericht und wissenschaftliche Abhandlungen*. Wiesbaden (Germany): Steiner.  
 PLOTT, ADALBERT 1963 *Bibliographie der Schriften Carl Ritters. Die Erde: Zeitschrift der Gesellschaft für Erdkunde zu Berlin* 94:13-36.  
 SCHMITTHENNER, HEINRICH 1937 Carl Ritter und Goethe. *Geographischer Zeitschrift* 43:161-175.  
 SCHMITTHENNER, HEINRICH 1951 *Studien über Carl Ritter*. Frankfurt geographische Hefte, Vol. 25, part 4. Frankfurt: Kramer.  
 SCHMITTHENNER, HEINRICH 1956 Carl Ritter. Volume 3, pages 189-200 in *Die grossen Deutschen: Deutsche Biographie*. Berlin: Propyläen-Verlag bei Ullstein.  
 WISOTZKI, EMIL 1897 *Zeitströmungen in der Geographie*. Leipzig: Duncker & Humblot.  
 WÜNSCHE, ALWIN 1899 *Die geschichtliche Bewegung und ihre geographische Bedingtheit bei Carl Ritter und bei seinen hervorragendsten Vorgängern in der Anthropo-geographie*. Dresden (Germany): Franke.

#### RITUAL

Citations in the *Oxford English Dictionary* from the fourteenth century on reveal two distinct trends of common usage for the words *rite* (*ritual*), *ceremony* (*ceremonial*), and *custom* (*customary*). On the one hand, these terms have been used interchangeably to denote any noninstinctive pre-

dictable action or series of actions that cannot be justified by a "rational" means-to-ends type of explanation. In this sense the English custom of shaking hands is a ritual, but the act of planting potatoes with a view to a harvest is not. But rationality is not easily defined. A psychiatrist may refer to the repeated hand washing of a compulsive neurotic as a "private ritual" even when, in the actor's judgment, the washing is a rational means to cleanliness. Likewise, a high-caste Hindu is required by his religion to engage in elaborate washing procedures to ensure his personal purity and cleanliness; the rationality or otherwise of such actions is a matter of cultural viewpoint. In this case, anthropologists who distinguish between ritual cleanliness and actual cleanliness are separating two aspects of a single state rather than two separate states. The distinction between cleanliness and dirt is itself a cultural derivation that presupposes an elaborate hierarchy of ritual values. If "nonrationality" is made a criterion of *ritual*, it must be remembered that the judge of what is rational is the observer, not the actor.

The other trend of usage has been to distinguish the three categories: ritual, ceremony, and custom. Ritual is then usually set apart as a body of custom specifically associated with religious performance, while ceremony and custom become residual categories for the description of secular activity. Where religion is the specific concern of fully institutionalized churches, as in Europe, a religious delimitation of ritual is unambiguous and easy to apply; in the exotic societies studied by anthropologists this is not the case. Recognizing this problem, some contemporary authors have argued that ambiguity in the data may be overcome by the multiplication of analytic concepts (e.g., Firth 1956, p. 46; Wilson 1954, p. 240; Gluckman 1962, pp. 20-24). Gluckman, in particular, favors an elaborate vocabulary giving clearly distinguishable meanings to *ceremony*, *ceremonious*, *ritual*, *ritualism*, and *ritualization*, but the circumstances in which precision might be useful are hard to imagine. Ritual is clearly not a fact of nature but a concept, and definitions of concepts should be operational; the merits of any particular formula will depend upon how the concept is being used.

In short, to understand the word *ritual* we must take note of the user's background and prejudices. A clergyman would probably assume that all ritual necessarily takes place inside a church in accordance with formally established rules and rubrics; a psychiatrist may be referring to the private compulsions of individual patients; an anthropologist will probably mean "a category of standardized

behavior (custom) in which the relationship between the means and the end is not 'intrinsic'" (Goody 1961, p. 159), but he will interpret this definition loosely or very precisely according to individual temperament. The associated terms *ceremonial* and *customary* are also used in very varied ways, even by professionals from the same discipline.

### Historical usage of the concept

The views of Robertson Smith (1889) are of particular relevance in arriving at a definition of terms. As a former professor of divinity, he advocated the study of comparative religion. He assumed that the boundaries between what is religion and what is not religion are self-evident. Modern religion (Christianity) consists of beliefs (dogma) and practices (ritual); "in the antique religions mythology takes the place of dogma." Myth is merely "an explanation of a religious usage." Hence, "the study of ancient religion must begin, not with myth, but with ritual and traditional usage." The thesis that religion consists essentially of beliefs and rituals and that, of the two, ritual is in some sense prior has influenced many later writers in many different fields.

Durkheim ([1912] 1954, p. 47) defined religion as "a unified system of beliefs and practices (rites) relative to sacred things, that is to say, things set apart and forbidden—beliefs and practices which unite into a single moral community called a Church, all those who adhere to them." Rites, for Durkheim, are "the rules of conduct which prescribe how a man should comport himself in the presence of these sacred objects" (*ibid.*, p. 41). Negative (ascetic) rites are the customs that we commonly label *taboo*. Positive rites include "imitative rites," which are in fact the same practices that Frazer called "homeopathic magic"; "representative or commemorative rites," which are the cults of ancestor worship; sacrifice; and placular rites, or memorials to misfortune, such as mourning. The overprecision of Durkheim's classification leads to some difficulties. He asserts dogmatically that "the division of the world into two domains, the one containing all that is sacred, the other all that is profane, is the distinctive trait of religious thought" (*ibid.*, p. 37). In his system, magic belongs to the sphere of the profane, even though "magic, too, is made up of beliefs and rites" (*ibid.*, p. 42).

The same set of rituals may readily be classified in other ways. Thus van Gennep (1908) proposed a category to cover all individual life-crisis ceremonials (e.g., those associated with birth, puberty,



marriage, death) and also recurrent calendric ceremonies such as birthdays and New Year's Day. He called these "rites of passage." In practice, van Gennep's schema has proved more useful than Durkheim's (see Gluckman 1962).

If Durkheim seems to be excessively rigid, Frazer, who was Robertson Smith's pupil, errs in the opposite direction. In the pages of *The Golden Bough* (1890) the words *custom*, *ceremonial*, *rite*, and *ritual* seem to be interchangeable. Belief and rite are assumed to be so closely interdependent that if evidence concerning either is available the author may confidently "conjecture" as to the other, which he does very freely. Employing similar assumptions, later writers have felt entitled to make the most sweeping reconstructions of ancient religious systems on the basis of slender archeological residues of ritual practice (Hooke 1958).

A more profitable development of Robertson Smith's theme was the inquiry by Jane Harrison (1912; 1913) into the relationship between ritual and art. Harrison noted that the Greek word *drama* is derived from *dromenon* (religious ritual, literally: "things done"). She attached special importance to Durkheim's category of "imitative rites": "Primitive man . . . tends to re-enact whatever makes him feel strongly; any one of his manifold occupations, hunting, fighting, later ploughing and sowing, provided it be of sufficient interest and importance, is material for a *dromenon* or rite." ([1913] 1951, p. 49). Thus ritual is seen as a magical dramatization of ordinary activities, while in turn the drama is a secular recapitulation of ritual. Although ritual is distinguished from non-ritual by the presence or absence of a religious context, the details of this distinction remain imprecise.

Harrison was a classical scholar who profited from the writings of anthropologists; the succeeding generation of anthropologists in turn profited from hers. Malinowski and Radcliffe-Brown introduced the concept of functionalism into British social anthropology; they were both indebted to Harrison, though both were also, and quite independently, the propagators of Durkheim's ideas. The concept of "ritual value," which Radcliffe-Brown developed in *The Andaman Islanders* (1922) and later writings, is essentially that espoused by Harrison. Objects to which ritual value attaches are objects that are socially important for secular reasons. Radcliffe-Brown, however, added the proposition that the performance of ritual generates in the actors certain "sentiments" that are advantageous to the society as a whole. In their discussions of this theme both Radcliffe-Brown and

Malinowski tend to assume that economic value depends upon utility rather than scarcity, and they attempt to distinguish ritual value as something other than economic value. Radcliffe-Brown shows that the Andamanese attached "ritual value" to objects (including foods) that were scarce luxuries, but he makes an unnecessary mystery of this fact. Karl Marx had a much clearer understanding of what is, after all, our common experience. Marx observed that the value of commodities in the market is quite different from the value of the same goods considered as objects of utility. He distinguishes the extra value that goods acquire by becoming market commodities as "fetichistic value" (1867-1879). This concept is closely akin to Radcliffe-Brown's "ritual value," though in Marx's argument the magical element is only an aspect of the commodity value, rather than the value as a whole. Furthermore, where Radcliffe-Brown urged that ritual is to the advantage of society, Marx claimed that it is to the disadvantage of the individual producer. The Marxist thesis is that in the activities of the secular market—where all values are supposed to be measured by the strictest canons of rationality—judgments are in fact influenced by mystical nonrational criteria. A full generation later Mauss (1925), developing his general theory of gift exchange from an entirely different viewpoint, reached an identical conclusion. Exchanges that appear to be grounded in secular, rational, utilitarian needs turn out to be compulsory acts of a ritual kind in which the objects exchanged are the vehicles of mystical power.

Of the authors I have mentioned, Durkheim, Harrison, Radcliffe-Brown, and Mauss all started out with the assumption that every social action belongs unambiguously to one or the other of two readily distinguishable categories: the nonrational, mystical, nonutilitarian, and sacred or the rational, common-sense, utilitarian, and profane. Each author would clearly like to distinguish a specific category, *ritual*, which could refer unambiguously and exhaustively to behavior relevant to things sacred. Each author ends up by demonstrating that no such discrimination is possible—that all "sacred things" are also, under certain conditions, "profane things," and vice versa. Malinowski sought to avoid this dilemma. For him, the essential issue was that of rationality rather than religion. Those who followed Frazer in thinking of magic as "a false scientific technique" necessarily classed magic as a profane activity, but according to Malinowski, primitive man has a clear understanding of the difference between a technical act

and a magical rite. Magic and religion both belong to a single sphere, the magico-religious; "... every ritual performance, from a piece of primitive Australian magic to a Corpus Christi procession, from an initiation ceremony to the Holy Mass, is a traditionally enacted miracle. . . . Man needs miracles not because he is benighted through primitive stupidity, [or] through the trickery of a priesthood . . . but because he realises at every stage of his development that the powers of his body and of his mind are limited" ([1923-1939] 1963, pp. 300-301).

But if ethnography offers little support to Durkheim, it offers still less to Malinowski. Most people in most societies have only the haziest ideas about the distinction either between sacred and profane or between rational and nonrational; it is a scholastic illusion to suppose that human actions are everywhere ordered to accord with such discriminations. Some authors still hold that a specific category is delimited by the phrase "behavior accounted for by mystical action": in my view they are mistaken.

### Ritual and communication

In this whole discussion two elements are involved which have so far been scarcely mentioned. Human actions can serve to *do* things, that is, alter the physical state of the world (as in lighting a bonfire), or they can serve to *say* things. Thus, the Englishman's handshake makes a complicated statement such as, "I am pleased to meet you and willing to converse." All speech is a form of customary behavior, but, likewise, all customary behavior is a form of speech, a mode of communicating information. In our dress, in our manners, even in our most trivial gestures we are constantly "making statements" that others can understand. For the most part these statements refer to human relationships and to status.

The actions that "say things" in this way are not as a rule intrinsically different from those that "do things." If I am cold, I am likely to put on more clothes, and this is a rational action to alter the state of the world; but the kinds of clothes I put on and the way I wear them will serve to "say things" about myself. Almost every human action that takes place in culturally defined surroundings is divisible in this way; it has a technical aspect which does something and an aesthetic, communicative aspect which says something.

In those types of behavior that are labeled ritual by any of the definitions so far discussed, the aesthetic, communicative aspect is particularly prominent, but a technical aspect is never entirely

absent. The devout Christian eats and drinks as part of a sacrament, but he also says grace as a preface to an ordinary meal. These are plainly "ritual" matters. But it is equally a matter of "ritual" that whereas an Englishman would ordinarily eat with a knife and fork, a Chinese would use chopsticks, and an Indian his right hand (but not his left, which for complex reasons is deemed polluted).

**The meaning of ritual.** Whether we use a narrow or a broad definition of ritual, one major problem is that of interpretation. What does ritual mean? If a ritual act be deemed to say something, how do we discover what it says? Clearly the actor's own view is inadequate. With minor variations the ritual of the Christian Mass is the same throughout Christendom, but each individual Christian will explain the performance by reference to his own sectarian doctrine. Such doctrines vary quite widely; the social scientist who seeks to understand why a particular ritual sequence possesses the content and form that he observes can expect little help from the rationalizations of the devout. But intuition is equally unreliable. Sacrifice, in the sense of the ritual killing of an animal victim, is an institution with a world-wide distribution. How can we explain this? Why should this particular kind of rite be considered an appropriate kind of action in the situations in which it is observed? There is no lack of theory. Some argue that the victim is identified with God and then sacramentally eaten; others that the victim is a gift or a bribe to the gods; others that the victim stands in substitution for the giver of the sacrifice; others that the victim is a symbolic representation of sin; and so on. All these explanations may be true or partly true for particular situations, but they cannot all be true at once, and none of them reach into the heart of the problem, which is, Why should the killing of an animal be endowed with sacramental quality at all?

Some interpretative approaches are more clearly formulated than others and deserve special attention. Radcliffe-Brown (1922) postulated that human beings always manipulate their thought categories consistently. We can discover what a ritual symbol means by observing the diverse uses of that symbol in both ritual and secular contexts. This is a powerful but by no means foolproof interpretative device. For example, the English speak of "high" status versus "low" status. We might then suppose that in ritual drama the person who is "higher" will always be superior. Up to a point this applies. Persons of authority are raised on a dais;



a suppliant kneels; an orator stands when his audience sits. But there are also situations where persons of extreme eminence sit (e.g., a king on his throne) when all others stand. The regularities are not simple.

This should not surprise us. In seeking to understand ritual we are, in effect, trying to discover the rules of grammar and syntax of an unknown language, and this is bound to be a very complicated business.

Lévi-Strauss (1962) is inclined to see ritual procedures as integral with processes of thought. The drama of ritual breaks up the continuum of visual experience into sets of categories with distinguishable names and thereby provides us with a conceptual apparatus for intellectual operations at an abstract and metaphysical level. Such an approach implies that we should think of ritual as a language in a quite literal sense. Various theorems of communication engineering and of structural linguistics should thus be applicable. We can, for example, start to investigate the role played by "redundancy" in ritual. Do binary contrasts in ritual correspond to phonemic contrasts in verbal speech forms? Can we discover, in any particular culture, rules concerning the development of a ritual sequence that would be comparable to the rules of generative grammar which Chomsky (1957) suggests must govern the modes by which each individual composes a verbal utterance? This is a field in which exploration has hardly begun.

**Ritual as social communication.** Most modern anthropologists would agree that culturally defined sets of behaviors can function as a language, but not all will accept my view that the term *ritual* is best used to denote this communicative aspect of behavior. Although we are still very much in the dark as to how ritual behaviors manage to convey messages, we understand roughly what the messages are about and at least part of what they say. Social anthropologists and sociologists alike claim that their special field is the study of systems of social relationship. This notion of social relationship is a verbal derivation based on inference. We do not observe relationships; we observe individuals behaving toward one another in customary, ritually standardized ways, and whatever we have to say about social relationships is, in the last analysis, an interpretation of these "ritual" acts. All of us in our private daily lives manipulate the symbols of an intricate behavioral code, and we readily decode the behavioral messages of our associates; this we take for granted. Comparable activities on a collective scale in the context of a religious institution are rated mysterious and irra-

tional. Yet their functional utility seems plain enough. Our day-to-day relationships depend upon a mutual knowledge and mutual acceptance of the fact that at any particular time any two individuals occupy different positions in a highly complex network of status relationships; ritual serves to reaffirm what these status differences are. It is characteristic of all kinds of ritual occasion that all participants adopt special forms of dress, which emphasize in an exaggerated way the formal social distinctions that separate one individual from another. Thus, ritual serves to remind the congregation just where each member stands in relation to every other and in relation to a larger system. It is necessary for our day-to-day affairs that we should have these occasional reminders, but it is also reassuring. It is this reassurance perhaps that explains why, in the absence of scientific medicine, ritual forms of therapy are often strikingly successful.

Here the argument seems to have come full circle. For if ritual be that aspect of customary behavior that "says things" rather than "does things" (cf. Parsons' instrumental-expressive dichotomy), how is it that, in the view of the actors (and even of some analysts), ritual may "do things" as well as "say things." The most obvious examples are healing rituals which form a vast class and have a world-wide distribution, but here we may also consider role-inversion rituals, which Gluckman (1962) has classed as "rituals of rebellion" and which he perceives as fulfilling a positive cathartic function.

**Ritual as power.** From the viewpoint of the actor, rites can alter the state of the world because they invoke power. If the power is treated as inherent in the rite itself, the analyst calls the action magic; if the power is believed to be external to the situation—a supernatural agency—the analyst says it is religious. Current argument on this theme is highly contentious, and I must declare my own position: I hold that the rite is prior to the explanatory belief. This will be recognized as essentially the view of Robertson Smith.

The concept of power itself is a derivation. We observe as an empirical fact that an individual *A* asserts dominance over another individual *B*; we observe that *B* submits to *A*, and we say that "A has power over B." And then in a ritual context we observe another individual *A'* going through a performance that he believes will coerce a fourth individual *B'*; or alternatively, we observe *B'* making a ritual act of submission to an unseen presence *C'*. The normal classification declares that the acts of *A* and *B* are rational but that the acts

of  $A^1$  and  $B^1$  are irrational. To me it seems that they are all actions of the same kind. The "authority" by which  $A$  is able to coerce and control the behavior of  $B$  in a secular situation is just as abstract and metaphysical as the magical power by which  $A^1$  seeks to coerce  $B^1$  or the religious power that  $B^1$  seeks to draw from  $C^1$ . Ideas about the relations between supernatural agencies and human beings or about the potency of particular ritual behaviors are modeled on first-hand experience of real-life relationships between real human beings. But conversely, every act by which one individual asserts his authority to curb or alter the behavior of another individual is an invocation of metaphysical force. The submissive response is an ideological reaction, and it is no more surprising that individuals should be influenced by magical performances or religious imprecations than that they should be influenced by the commands of authority. The power of ritual is just as actual as the power of command.

**Ritual as belief.** Unlike Robertson Smith, Tylor (1871) assumed the priority of belief over ritual. In England a Neo-Tylorian view has a number of contemporary advocates. According to their argument, it is the belief accompanying the behavior, rather than any quality of the behavior itself, that distinguishes ritual. Since the participants in a religious ritual claim that their actions are designed to alter the state of the world by bringing coercive influence upon supernatural agencies, why should we not accept this statement at its face value? Why invoke the proposition that the actions in question are, as Durkheim would have it, "symbolic representations of social relationships." Goody counterposes the intellectualized interpretation of social behavior made by the observer to the statement of the actor himself. "What happens, then," writes Goody (1961, p. 157), "is that symbolic acts are defined in opposition to rational acts and constitute a residual category to which 'meaning' is assigned by the observer in order to make sense of [the] otherwise irrational. . . ." Ritual acts are to be interpreted in the context of belief: they mean what the actors say they mean. This common-sense approach clearly has its attractions. Yet it may be argued that if culturally-defined behavior can only be interpreted by the actors, all cross-cultural generalization is impossible, and all attempts to make a rational analysis of the irrational must necessarily be fallacious (see Goody 1961, p. 155). In contrast, I, along with other Durkheimians, continue to insist that religious behavior cannot be based upon an illusion.

Jane Harrison's thesis that "ritual is a dramatisation of myth" was reformulated by Malinowski in the assertion that "myth is a charter for social action." According to this argument, myth and ritual are not merely interdependent; they jointly provide a model for "correct" moral attitudes in secular life. But although it is easy to cite examples in which rituals enshrine in a quite straightforward way the most strongly felt values of society, there are many striking exceptions. The characters of myth frequently break all the moral conventions of mundane society in the most glaring way, and in many rituals the actors are required to behave in a manner precisely contrary to that which they would be expected to adopt in ordinary life. Two very different types of explanation have been offered for facts of this kind. One sees this role inversion as symbolic; the events of myth and ritual refer to the space-time of "the other world"; they belong to Durkheim's category of the *sacred*, and to express this fact their content systematically inverts whatever is appropriate to "this world," the *profane*. In contrast, Gluckman, in an argument that has wide application (see Norbeck 1961, pp. 205-211), stresses the aggressive elements present in role-reversal ceremonies, which he aptly names "rituals of rebellion" (Gluckman 1962). The performers, he suggests, act out in dramatic form hostilities that are deeply felt but may not be expressed in normal secular relationships. This acted aggression serves as a cathartic release mechanism, and by relieving tension these inverted behaviors actually serve to strengthen the moral code they appear to deny. It is an ingenious argument but hard to validate. Once again we are faced with the difficulty that sharply contrasted interpretations seem to afford partial explanations of the same ethnographic facts, so that choice of theory becomes a matter of personal predilection.

Nineteenth-century positivist thinkers made a triadic distinction between reason, magic, and religion. Various authors have attempted to fit ritual to this triad and also to the two dichotomies: sacred-profane, rational-nonrational. Some of the resulting difficulties have been considered. It is argued that no useful distinction may be made between ritual acts and customary acts but that in discussing ritual we are concerned with aspects of behavior that are expressive (aesthetic) rather than instrumental (technical). Ritual action, thus conceived, serves to express the status of the actor vis-à-vis his environment, both physical and social; it may also serve to alter the status of the actor.



When ritual functions in this latter sense, it is a manifestation of power; thus, the universal belief in the potency of ritual action is by no means an illusion. No attempt has been made to discuss the forms of ritual. Any form of secular activity, whether practical or recreational, can be stylized into dramatic performance and made the focus of a ritual sequence. Such stylization tends to distort the secular norm in either of two directions: the emphasis may be ascetic, representing the intensification of formal restraint, or ecstatic, signifying the elimination of restraint. Ascetic and ecstatic elements are present in most ceremonial sequences, and the contrast may form part of the communication code (Leach 1961, chapter 6). Finally, it has been stressed that even among those who have specialized in this field there is the widest possible disagreement as to how the word ritual should be used and how the performance of ritual should be understood.

EDMUND R. LEACH

[Directly related are the entries COMMUNICATION; INTERACTION, article on DRAMATISM; MYTH AND SYMBOL; POLLUTION; and the biographies of DURKHEIM; FRAZER; MALINOWSKI; RADCLIFFE-BROWN; SMITH, WILLIAM ROBERTSON.]

#### BIBLIOGRAPHY

- CHOMSKY, NOAM (1957) 1964 *Syntactic Structures*. The Hague: Mouton.
- DURKHEIM, ÉMILE (1912) 1954 *The Elementary Forms of the Religious Life*. London: Allen & Unwin; New York: Macmillan. → First published as *Les formes élémentaires de la vie religieuse, le système totémique en Australie*. A paperback edition was published in 1961 by Collier.
- FIRTH, RAYMOND 1956 Ceremonies for Children and Social Frequency in Tikopia. *Oceania* 27:12-55.
- FRAZER, JAMES (1890) 1955 *The Golden Bough*. 13 vols., 3d ed., rev. & enl. New York: St. Martins; London: Macmillan.
- GENNEP, ARNOLD VAN (1909) 1960 *The Rites of Passage*. London: Routledge; Univ. of Chicago Press. → First published in French.
- GLUCKMAN, MAX 1962 *Les rites de passage*. Pages 1-52 in Max Gluckman (editor), *Essays on the Ritual of Social Relations*. Manchester (England) Univ. Press.
- GOODY, J. R. 1961 Religion and Ritual: The Definitional Problem. *British Journal of Sociology* 12:142-164.
- HARRISON, JANE E. (1912) 1937 *Themis: A Study of the Social Origins of Greek Religion*. 2d ed., rev. Cambridge Univ. Press.
- HARRISON, JANE E. (1913) 1951 *Ancient Art and Ritual*. Rev. ed. New York: Oxford Univ. Press.
- HOOKE, SAMUEL H. (editor) 1958 *Myth, Ritual, and Kingship*. Oxford: Clarendon.
- LEACH, EDMUND R. 1961 *Rethinking Anthropology*. London School of Economics and Political Science, Monographs on Social Anthropology, No. 22. London: Athlone.

- LÉVI-STRAUSS, CLAUDE (1962) 1966 *The Savage Mind*. Univ. of Chicago Press. → First published in French.
- MALINOWSKI, BRONISLAW (1913-1941) 1962 *Sex, Culture and Myth*. New York: Harcourt. → See especially pages 295-336, "The Foundations of Faith and Morals."
- MARX, KARL (1867-1879) 1925-1926 *Capital: A Critique of Political Economy*. 3 vols. Chicago: Kerr. → Volume 1: *The Process of Capitalist Production*. Volume 2: *The Process of Circulation of Capital*. Volume 3: *The Process of Capitalist Production as a Whole*. The manuscripts of Volumes 2 and 3 were written between 1867 and 1879. They were first published posthumously in German in 1885 and 1894.
- MAUSS, MARCEL (1925) 1954 *The Gift: Forms and Functions of Exchange in Archaic Societies*. Glencoe, Ill.: Free Press. → First published as *Essai sur le don: Forme et raison de l'échange dans les sociétés archaïques*.
- NORRBECK, EDWARD 1961 *Religion in Primitive Society*. New York: Harper.
- RADCLIFFE-BROWN, A. R. (1922) 1948 *The Andaman Islanders*. Glencoe, Ill.: Free Press.
- SMITH, WILLIAM ROBERTSON (1889) 1956 *The Religion of the Semites*. New York: Meridian.
- SRINIVAS, MYSORE N. 1952 *Religion and Society Among the Coorgs of South India*. Oxford: Clarendon.
- TYLOR, EDWARD B. (1871) 1958 *Primitive Culture: Researches into the Development of Mythology, Philosophy, Religion, Art and Custom*. 2 vols. Gloucester, Mass.: Smith. → Volume 1: *Origins of Culture*. Volume 2: *Religion in Primitive Culture*.
- WILSON, MONICA 1954 *Nyakyusa Ritual and Symbolism*. *American Anthropologist* New Series 56:228-241.

#### RIVERS, W. H. R.

William Halse Rivers (1864-1922), British anthropologist and medical psychologist, was educated at Tonbridge School and St. Bartholomew's Hospital, London. He held a medical degree, was elected a fellow of the Royal College of Physicians, and was a lecturer in experimental psychology at Guy's Hospital, London. From 1897 on, he was a lecturer at Cambridge University and was responsible for founding the Cambridge school of experimental psychology. He became a fellow of St. John's College, Cambridge, in 1902. In 1898 he joined the Cambridge anthropological expedition to Torres Strait led by A. C. Haddon and was in charge of the psychological work. He made several ethnological investigations: of the Todas in south India in 1902, and in Melanesia, in 1908 and again in 1914. During World War I Rivers performed distinguished service as a psychopathologist working in military hospitals. He was one of the few English medical men of that period to recognize the far-reaching significance of the work of Freud; he himself made important contributions to the study of war neurosis. He received honorary

degrees from the universities of St. Andrews and Manchester and was a fellow of the Royal Society. He was president of the Folklore Society in 1920–1921 and president of the Royal Anthropological Institute in 1921–1922. His death following an operation was unexpected; he was only 58 and actively engaged in many intellectual activities.

**Anthropological career.** Rivers' lifelong interest in the twin disciplines of medical psychology and anthropology had long antecedents. His father, the Reverend H. F. Rivers, had studied under his own brother-in-law, James Hunt, the owner of a fashionable (and very profitable) institute of speech therapy, which had been founded by Hunt's father early in the century. On Hunt's death in 1869, H. F. Rivers took over the institute and brought out a revised edition of Hunt's book, *Stammering and Stuttering: Their Nature and Treatment*. But Hunt, besides being a practical exponent of experimental psychology, had been very prominent in the early history of British anthropology. He had been honorary secretary of the Ethnological Society of London from 1859 to 1862 and then founded and financed a breakaway institution called the Anthropological Society of London, which survived as a separate entity until 1871. The present Royal Anthropological Institute, of which Rivers was president at the time of his death, is a successor institution to the two parent bodies with which his maternal uncle had been associated at the time of Rivers' birth.

Rivers' professional status was that of a psychologist, but he is principally remembered for his contributions to anthropology. On the Torres Strait expedition of 1898 he developed a keen interest in ethnographic problems, particularly in the algebraic peculiarities of kin-term systems. He devised a "genealogical method" for recording the mutual kinship connections of the members of a closed community, a method that proved to be of lasting importance (Rivers 1900). Rivers brought to his anthropological studies a scientific detachment and experimental rigor wholly different from the spirit of classical scholarship that pervades the writings of his celebrated Cambridge colleague Sir James Frazer. It is primarily because of Rivers that most British social anthropologists now think of themselves as being engaged in a science rather than a literary exercise. By the standards of its time *The Todas* (1906) was an outstanding example of precise documentation, and for many years this book served British anthropologists as a model for ethnographic monographs.

Like all his contemporaries, Rivers took it for granted that the objective of the anthropologist is

to reconstruct the history of the primitive peoples whom he studies. During the early part of his career Rivers' historical assumptions were those of an evolutionist. In this respect he was a follower of L. H. Morgan, whose studies of kinship terminologies (Morgan 1871) had been largely neglected in England. The functionalist reaction against evolutionism, under the twin but discordant trumpets of B. Malinowski and A. R. Radcliffe-Brown, did not achieve general acceptance until some years after Rivers' death. Even so, Radcliffe-Brown, who made his expedition to the Andaman Islands in 1906 under the tutelage of Haddon and Rivers, retained throughout his career a strong interest in the problems of kinship terminology that had so obsessed Rivers. Although this interest led Rivers into many fanciful excesses of "kinship algebra," whereby he deduced sociological conditions that can certainly never have existed, it was a bias which proved to be a crucial step in the development of the "structuralist" orientation of contemporary British social anthropology (which derives from Rivers by way of Radcliffe-Brown) as distinct from the "cultural" orientation which dominated the work of Tylor, Frazer, and Malinowski.

Around 1910 Rivers developed a close friendship with Sir Grafton Elliot Smith, a medical colleague who advocated an extreme diffusionist doctrine, claiming that the whole of human civilization was derived from that of ancient Egypt [see DIFFUSION, article on CULTURAL DIFFUSION]. Rivers' uncritical acceptance of Smith's dogmatic and wildly exaggerated assertions is a matter for astonishment. Nearly all of Rivers' later anthropological writings, including the very important *History of Melanesian Society* (1914a) and *Kinship and Social Organization* (1914b), show Smith's influence to a marked degree, usually to their detriment. Rivers' anthropological textbook, *Social Organization* (1924b), although it was mutilated by its editors, Elliot Smith and W. G. Perry, and was certainly not in the same class as R. H. Lowie's *Primitive Society* (1920), still merits serious attention.

Rivers' discrimination of concepts has a scientific clarity that is commonly lacking in anthropological writing. Although the problems of *descent*, *succession*, and *inheritance* are not so simple as Rivers supposed, he at least perceived the nature of problems that many later writers have ignored. Among his earlier works the ethnographic accounts of the Todas (1906) and of the Banks Islanders (1914a, vol. 1) have permanent merit, but the high esteem formerly placed upon his theoretical studies of kinship terminologies now seems misplaced. In studying kin terms, Rivers, like Morgan,



assumed that classificatory usages must be "survivals" of an earlier state of society. Rivers' most astonishing historical reconstruction of this kind occurs in *Kinship and Social Organisation* (1914b, pp. 33-38), where he invented a social system that required a man to marry either his mother's brother's wife or his brother's daughter's daughter. The reconstructions of history found in more modern work, such as Murdock's *Social Structure* (1949), lack the preposterous quality which marks many of Rivers' examples, but they rest upon assumptions concerning the nature of kinship that are basically very similar.

In sum, very little of Rivers' "kinship theory" can still be taken seriously, but its influence can still be traced. In England, Radcliffe-Brown, Brenda Seligman, J. Layard, and A. M. Hocart all produced work in this field which derived from that of Rivers and which has had a lasting influence upon the thinking of British social anthropologists, even though the latter no longer concern themselves with historical reconstructions. Oddly enough, although Rivers' influence in the United States is very indirect, there are some contemporary American writers, such as Robert and Barbara Lane, who stand very close indeed to Rivers' way of thinking. Rivers' influence on his psychological successors was much less pronounced, though some early work by his friend F. C. Bartlett, who was later professor of experimental psychology at Cambridge, perpetuates Rivers' ideas (e.g., Bartlett 1923).

EDMUND R. LEACH

[For the historical context of Rivers' work, see the biographies of HADDON and MORGAN, L. H. For discussion of the subsequent development of his ideas, see SOCIAL STRUCTURE; and the biographies of BARTLETT and RADCLIFFE-BROWN.]

#### WORKS BY RIVERS

- 1900 A Genealogical Method of Collecting Social and Vital Statistics. *Journal of the Royal Anthropological Institute of Great Britain and Ireland* 30:74-82.  
 1906 *The Todas*. New York and London: Macmillan.  
 1908 *The Influence of Alcohol and Other Drugs on Fatigue*. London: Arnold.  
 1914a *The History of Melanesian Society*. 2 vols. Percy Sladen Trust Expedition to Melanesia, Publication No. 1. Cambridge Univ. Press.  
 1914b *Kinship and Social Organisation*. London School of Economics and Political Science, Studies, No. 36. London: Constable.  
 (1920) 1922 *Instinct and the Unconscious*. 2d ed. Cambridge Univ. Press.  
 1923a *Conflict and Dream*. London: Routledge; New York: Harcourt.  
 1923b *Psychology and Politics, and Other Essays*. London: Routledge; New York: Harcourt.

- 1924a *Medicine, Magic, and Religion*. London: Routledge; New York: Harcourt.  
 1924b *Social Organization*. London: Routledge; New York: Knopf.  
 1926 *Psychology and Ethnology*. London: Routledge; New York: Harcourt.

#### SUPPLEMENTARY BIBLIOGRAPHY

- BARTLETT, F. C. 1923 *Psychology and Primitive Culture*. Cambridge Univ. Press; New York: Macmillan.  
 HADDON, ALFRED C.; and BARTLETT, F. C. 1922 William Halse Rivers Rivers, M.D., F.R.S. With a Bibliography Compiled by Ethel S. Fegan. *Man* 22:97-107.  
 LOWIE, ROBERT H. (1920) 1947 *Primitive Society*. New York: Liveright. → A paperback edition was published in 1961 by Harper.  
 MORGAN, LEWIS H. 1871 *Systems of Consanguinity and Affinity of the Human Family*. Smithsonian Contributions to Knowledge, Vol. 17; Publication No. 218. Washington: Smithsonian Institution  
 MURDOCK, GEORGE P. 1949 *Social Structure*. New York: Macmillan.

#### RIVET, PAUL

Paul Rivet, French ethnologist and Americanist, was born at Wassigny, in the Ardennes, in 1876 and died in Paris in 1958. He studied medicine at the Faculty of Medicine and the École du Service de Santé Militaire of Lyons, where he obtained the degree of doctor of medicine in 1898. In 1901 he was attached to the scientific mission of General Bourgeois, which was sent to Ecuador to verify certain terrestrial measurements made at the end of the eighteenth century by Pierre Bouguer, Louis Godin, and Charles Marie de La Condamine. After the mission returned, Rivet stayed on for six years, studying the populations of the high valleys of the Andes. From that time on, he devoted himself to anthropological and linguistic researches.

When Rivet returned to France, the Musée National d'Histoire Naturelle engaged him to classify and study the collections of anthropological and linguistic documents that he brought back (1912-1922). He became an assistant to René Verneau, who held the chair of anthropology at the museum, and then assistant director of the laboratory, a post he held until he was himself appointed to the chair of anthropology in 1928. The focus of his interest was the geology, geography, and linguistics of the Americas. He organized weekly meetings at his home, at which both French and foreign scientists presented and discussed the results of their work on the Americas; these meetings continued for fifty years, interrupted only by two world wars.

In 1926 Rivet took part in creating the Institut d'Ethnologie of the University of Paris, where he

helped to train a great many French and foreign ethnologists. He took over the chair of anthropology of the Musée National d'Histoire Naturelle in 1928 and later succeeded in having one wing of the new Palais de Chaillot allotted to an ethnographical museum, eventually transforming this museum into the great Musée de l'Homme, famous for its fossil-man collections. In 1942 he was welcomed to Colombia by President Eduardo Santos and there founded the Colombian Institute of Ethnology and set up a Musée de l'Homme in Bogotá that was modeled on the one in Paris. After a period in Mexico as cultural delegate of Free France, Rivet resumed, in 1945, his chairs at the museum and the Institut d'Ethnologie, as well as the directorship of the Musée de l'Homme. He also continued his research in South America and sparked the activities of the Société des Américanistes in Paris.

The basic concept of Rivet's extensive work is the notion that there is "a real interdependence, however hard it may sometimes be to prove it, among linguistics, ethnology, and [physical] anthropology" (1927, p. 14). In tracing the origins of American aboriginals, therefore, he compared anthropometric studies of fossils with those made on living peoples; he also studied distributions of cultural elements, such as language, religion, music, games, weapons, household furnishings, and tools, as well as of blood groups and pathologies. Combining these studies with the results of research done by others, he developed his central thesis—that the peopling of America did not derive exclusively from Asia. In opposition to Hrdlička's theory that the population of America came only from Asia, via the Bering Strait, the Aleutians, and the northwest, a theory that at that time still had considerable support, Rivet believed that two other sources also contributed to the peopling of South America. He did not deny that the major part of the population had come from Asia but insisted that there had been migrations (1) from Australia, dating back at least six thousand years, and (2) from Melanesia, at a later but still ancient period. The migration from Australia must have followed the shores of the Antarctic, making use of the chains of islands that bind that continent to Australia on the one hand and to America on the other ([1943] 1957, p. 24; 1956). The Melanesian migration, he believed, took place in a succession of waves, by the maritime route of the South Pacific. In his publications on this subject, Rivet presented not only the anthropological, linguistic, and other evidence for his theory about these migrations but also described the similarities between the primi-

tive populations of South America and Oceania that initially led him to formulate it.

In his linguistic work he pointed out many relationships of a fairly narrow scope, almost always with great reliability. His contribution in providing considerable data on otherwise unknown languages was also considerable.

RAYMOND RONZE

[For the historical context of Rivet's work, see the biography of HRDLÍČKA. For discussion of the subsequent development of his ideas, see DIFFUSION, article on CULTURAL DIFFUSION.]

#### WORKS BY RIVET

- 1912 Entente internationale pour l'unification des mesures anthropométriques sur le vivant. *Anthropologie* (Paris) 23:623-627.
- 1912-1922 VERNEAU, RENÉ; and RIVET, PAUL Ethnographie ancienne de l'Équateur. 2 parts. France, Service Géographique de l'Armée, *Mission . . . pour la mesure d'un arc de méridien équatorial en Amérique du Sud . . . 1889-1906* 6: parts 1-2.
- 1924 Les Mélanésio-Polynésiens et les Australiens en Amérique. Académie des Inscriptions et Belles-Lettres, Paris, *Comptes rendus des séances* [1924]: 335-342.
- 1925 Les origines de l'homme américain. *Anthropologie* (Paris) 35:293-319.
- 1926 Le peuplement de l'Amérique précolombienne. *Scientia: Rivista di scienza* 40:89-100.
- 1927 Titres et travaux scientifiques de P. Rivet. Mâcon (France): Protat.
- (1943) 1957 *Les origines de l'homme américain*. Paris: Gallimard.
- 1956 Early Contacts Between Polynesia and America. *Diogenes* 16:78-92.

#### SUPPLEMENTARY BIBLIOGRAPHY

- CARTER, GEORGE F. 1957 The American Civilization Puzzle. *Johns Hopkins Magazine* 8, no. 5:8-11, 21-24.
- HRDLÍČKA, ALEŠ 1917a The Genesis of the American Indian. Pages 559-568 in International Congress of Americanists, Nineteenth, Washington, 1915, *Proceedings*. Washington: The Congress.
- HRDLÍČKA, ALEŠ 1917b The Old White American. Pages 582-601 in International Congress of Americanists, Nineteenth, Washington, 1915, *Proceedings*. Washington: The Congress.

#### ROBERTSON, DENNIS HOLME

Sir Dennis Holme Robertson was born at Lowestoft, England, in 1890 and died at Cambridge in 1963. He made his most important contribution to economics in the fields of monetary theory and industrial fluctuation, but his range was wide, and his writings cover many aspects of economics and the economic and social problems of his time. Like many of his predecessors and contemporaries, he came to economics from another discipline. In 1908 he entered Trinity College, Cambridge, with



a major scholarship in classics, and two years later he was placed in the first division of the first class of the classical tripos and won the Craven scholarship. In three successive years he also won the chancellor's medal for English verse, and he continued to write poetry for most of his life. He had a superb mastery of the language, and his writing is well known for its clarity, conciseness, and elegance. In 1910 he abandoned classics for economics. At this time economics at Cambridge was dominated by the work of Alfred Marshall, which remained the principal influence on Robertson. Robertson became a pupil of J. M. Keynes (his senior by seven years), thus beginning a collaboration that was to last for more than twenty years. He obtained a first class in the economics tripos in 1912 and turned to his first piece of research, which was completed early in 1914 and published in 1915 as *A Study of Industrial Fluctuation*.

**Work on the trade cycle.** *A Study of Industrial Fluctuation* is a remarkable work for so young an economist. Robertson collected a mass of empirical data, and although some of his methods appear amateurish by later standards (he was working singlehandedly and without any statistical aids), he derived conclusions of great importance. In particular, he stressed the role of invention and innovation in determining the course of investment and therefore of general economic activity, thus helping to explain the causation and mechanics of the trade cycle. This new approach also implied that fluctuations are closely associated with economic progress and are to some extent part of the price to be paid for it—a conclusion which is not always now accepted but which Robertson continued to hold. This is not to say that he considered that countercyclical policy need be totally ineffective: this early work contains suggestions which were well in advance of their time, such as “a somewhat saner and more centralized investment policy” ([1915] 1948, p. 246) and an attack on what in the 1930s became known as the “Treasury view” of the ineffectiveness of public expenditure as a means of combating depressions.

Apart from its particular conclusions, three features of the book were to remain characteristic of all his later work. First, there is very little in it about “monetary” problems as such. In a new introduction to the 1948 reprint, he stated that he had “been taught to be constantly trying to dig down below the money surface of things” (p. xii), and even in his later work in monetary theory he constantly turned to the influence of “real” forces. Thus, for example, he continued to emphasize the importance of the classical forces of productivity

and thrift as factors determining the rate of interest, unlike economists who regarded it as a purely “monetary” phenomenon determined in the market for bonds. Second, there is an awareness of the complexity of the phenomenon of fluctuation and of all economic phenomena, which left him “with an abiding sense of the difficulty of providing . . . neat little models of the trade cycle and (*a fortiori*) neat little packets of therapeutic pills” (p. x). Third, there is emphasis on structural problems. Although the theme is fluctuation in real income as a whole (in itself a new emphasis, which was rediscovered by some writers in the 1930s), the treatment is permeated by recognition of the effects of changes in one sector of the economy, or in one industry, on the other sectors and industries, and therefore on the aggregate.

**Early work on monetary theory.** When *A Study of Industrial Fluctuation* was published in 1915, Robertson was in the army. He had joined up on the first day of World War I and served throughout, mainly in Egypt and Palestine, as a transport officer. After the war he returned to Cambridge, where he stayed until 1938. It was during this period that he turned to intensive work in monetary theory, at first in close collaboration with Keynes but later as one of his foremost critics. His contributions are contained mainly in two small books, *Money*, published in 1922 (with revised editions in 1924 and 1928 and many subsequent reprints), and *Banking Policy and the Price Level*, published in 1926. The former is a general textbook in a series edited by Keynes, but it is also an important contribution to the development of ideas. Robertson's aim was to “try to re-integrate the theory of money into that of the trade cycle” ([1915] 1948, p. xv). The period in which he wrote was one of monetary instability; the inflation of the war and postwar years was followed by collapse. Traditional financial policies were inadequate to deal with instability and even partly contributed to it, and the mechanisms involved were very imperfectly understood. Virtually the only theory was a rather crude form of quantity theory, which attributed rising prices to the effects of increasing note issue. Robertson's main innovation (and it was fundamental) was to shift the emphasis from the issue of notes to the creation of bank credit and then to break down the process of inflation into a sequence of stages; he also considered the reverse process of deflation (this was in 1922 when the fall in prices had started), giving a warning against too precipitate a return to the gold standard. This analysis was developed in *Banking Policy and the Price Level*, as well as in the later

editions of *Money*, its central theme being the key importance of saving and investment as the determinants of the level of economic activity, with the banking system as the means whereby savings are translated (or not translated) into investment. The step-by-step method of analysis is carried further here than in the earlier work. The implications for policy were the same as those he explicitly stated later: "... the ideal banking policy might be one which was founded on the principle of price-stabilisation as a norm, but which was ready to see the fruits of a prolonged and general increase in individual productivity shared in the form of lower prices, and perhaps to acquiesce in moderate price-rises in order that advantage might be taken of discontinuous leaps in industrial technique" ("Theories of Banking Policy" in [1924-1940] 1956, p. 59). He consistently maintained this view; it is found, for example, in his last work, *Memorandum Submitted to the Canadian Royal Commission on Banking and Finance* (1963).

**Problems of industrialism.** In 1923, immediately after writing *Money*, Robertson published *The Control of Industry*, another volume in the series of textbooks edited by Keynes. The theme is the problem of "how, if at all, [we can] ensure that the men and women engaged in industry shall not become mere instruments of production or mere passive receptacles of its fruits, but shall retain, in their relation to the economic circumstances of their life, the character of self-directing human beings?" ([1923] 1960, pp. 1-2).

It was characteristic of Robertson not to offer easy solutions but rather to present a penetrating and balanced discussion of the nature of the problems to be solved. The scope of the book is, indeed, wider than this theme would suggest, for in its short compass it is no less than a critique of the underlying principles and typical organizations of modern capitalism. He identified the following tenet as the main principle of free enterprise: "Where the risk lies, there the control lies also." He explored its application to a wide range of proposals and attempts to change the methods of control of industry, ranging from revolutionary syndicalism to consumers' cooperatives and workers' copartnership. Although he did not claim any originality of thought on the subject, the work was a minor classic and is still widely used.

**Response to Keynes's work.** In the introduction to *Banking Policy and the Price Level*, Robertson stated that his collaboration with Keynes had been so close that "neither of us now knows how much of the ideas . . . is his and how much is mine" ([1926] 1949, p. 5). With the publication of

Keynes's *General Theory of Employment, Interest and Money* in 1936, however, their paths diverged. Robertson's criticisms of the Keynesian system are contained in a series of articles, most of which were later republished in volumes of collected essays (see 1924-1940; 1952; 1956). Many of his criticisms were regarded in some quarters as carping at minutiae of a fundamental revolution in economic thinking that he had failed to understand. To Robertson, however, much of the Keynesian "revolution" was retrogression rather than advance. So far as the main analysis was concerned, Robertson regarded Keynes's treatment of the savings-investment complex in terms of static and stable equilibrium as a step backward, leading to confusion and misunderstanding. Although much of the controversy seems to have been about words, especially the definitions of saving and investment, the issue of whether savings and investment are or are not equal in equilibrium was a fundamental one. Similarly crucial was what was meant by equilibrium: for Keynes it was short-period equilibrium; for Robertson it was equilibrium over time, to be analyzed and understood in detail only by his step-by-step method. Further, Robertson's "equilibrium" also implied (indeed, took for granted) the growth of the economy in terms of measurable real output or income. Keynes, on the other hand, was led to the "stagnation thesis," which Robertson regarded as unwarranted analytically and also as being a generalization from one severe but atypical depression. Moreover, the implications for policy of Keynes's system were unambiguously inflationary and thus a departure from Robertson's ideal of ordered progress and monetary stability.

In retrospect, much of the controversy over the *General Theory* seems to have been unduly labored, in part perhaps as a result of the defects of the book and of the extreme positions taken by some of the participants. The Keynesian system was considerably modified and developed under criticism, not least that of Robertson, and the accepted view on some matters became as much Robertsonian as Keynesian. For example, Robertson was right in pointing out that the equality of savings and investment was treated by many of Keynes's followers both as an identity and as an equality achieved by movement of income to an equilibrium level, and many of his further criticisms flowed from this. The controversy over the determinants of the rate of interest showed that a modified Keynesian "liquidity preference" theory and the Robertsonian "loanable funds" approach could in certain circumstances give much the same results, and some economists regard the Robertsonian formulation as



the more fruitful in interpreting dynamic situations. And although Robertson remained out of sympathy with the Keynesian system, his own views underwent modification as he reconsidered his earlier ideas, many of which had originally been worked out in collaboration with Keynes: for example, his discussions after 1945 of problems of inflation would have been very different had it not been for the *General Theory* and its aftermath.

**Later work.** In 1938 Robertson left Cambridge to take up a professorship at the London School of Economics and at the outbreak of World War II joined the Treasury, where his main concern was with the balance of payments and overseas finance. In 1944 he was one of the team accompanying Keynes to the Bretton Woods conference, and he played an important part in the negotiations leading to the final agreement. At the end of the year he returned to Cambridge to succeed A. C. Pigou in the chair of political economy. In that capacity he lectured on economic principles until his retirement in 1957. His lectures, published in three volumes from 1957 to 1959, were in the main an exposition of the problems of value and distribution (although Volume 3, covering only a dozen lectures given in the short third term at Cambridge, is concerned with money and fluctuations). Almost inevitably he based himself on Marshall, but he had kept up with the flood of literature before and after the war, and the lectures constitute an important statement of the present state of some of the central and traditional problems of economics, with much penetrating comment, presented in his own inimitable style. In these years he also wrote a large number of papers on current problems of domestic and international economic policy, as well as some on more theoretical issues; most of these were collected into published volumes of essays (e.g., 1956; 1957-1959). It was at this time that formal worldwide recognition came; he had received an honorary degree from Harvard University at the tercentenary celebration in 1936 and from the University of Louvain, Belgium, in 1947; then, in quick succession, he was honored by the universities of Durham, Manchester, Sheffield, Amsterdam, London, and Columbia. He was knighted in 1953, in recognition of his standing as an economist rather than as a public servant (he had first received public honors on leaving the Treasury in 1944).

Robertson did take some part in public affairs, although usually with some reluctance and out of a strong sense of duty. In particular, from 1944 to 1946 he was a member of the Royal Commission on Equal Pay, exerting a powerful influence on its report. In 1957-1958 he was a member of

the three-man Council on Prices, Productivity and Incomes set up by the chancellor of the Exchequer "to keep under review changes in prices, productivity and the level of incomes (including wages, salaries and profits) and to report thereon from time to time." He was the only economist member and undertook the task with particular reluctance; he had serious doubts about the whole matter of "incomes policy" with which the council was linked, and he knew that his own ideas, which he would be bound to put forward, would be generally unwelcome in the current climate of political and economic opinion. He proved to be right: the trade unions refused to have anything to do with the council, and its first and principal report, in February 1958, had a mixed reception. Although some commentators regarded it as an important contribution to thinking about the current situation (and it contains a good deal of hard analysis behind its apparent simplicity), it was attacked mostly for the very reasons that Robertson had expected, such as its emphasis on the high level of demand, supported by a plentiful supply of money and the pursuit by governments of "full employment policies," rather than on the wage claims of unions, as the main cause of inflation; certain comments on incomes policy and wage restraint; and a Robertsonian preference for monetary stability and falling prices as a result of increasing productivity, rather than for rising prices as a continuous stimulus to activity. After publication of a second report in August 1958, which brought the statistics up-to-date and suggested further investigations, Robertson resigned. His two colleagues resigned in 1960, and a reconstituted council produced one further report before falling into desuetude.

In 1960 Robertson was invited to give the Marshall lectures at Cambridge and chose as his theme *Growth, Wages, Money* (1961)—a complex of subjects that were highly topical but that also took him back to his first work on industrial fluctuation. In his opening statement, he remarked "... I have sometimes hoped that I could be useful by taking some theme which was really too difficult for me and over which the giants were in conflict ... and giving, for the benefit of those I called my fellow-idiot, my impression of the state of the debate" (p. 4). He not only recorded the state of the debate but made many penetrating comments; he quietly pricked some of the more pretentious bubbles of economic theory and policy, and displayed a salutary skepticism about the value of growth "models," of "neat little models of the trade cycle," and of incomes policy as means of dealing with inflation and of ensuring growth. Although he cast doubt

## ROBINSON, JAMES HARVEY

on many of the economic shibboleths of the time, he would have hoped not to "be debarred from being classed among those who genuinely desire to find means to limit the turbulence, without destroying the vitality, of the 'procreant urge of the world'" ([1915] 1948, p. xvii).

S. R. DENNISON

[For the historical context of Robertson's work, see the biographies of KEYNES, JOHN MAYNARD, and MARSHALL. For discussion of the subsequent development of his ideas, see BANKING, CENTRAL; BUSINESS CYCLES; INCOME AND EMPLOYMENT THEORY; MONETARY POLICY; MONEY. Also related is the biography of PIGOU.]

## WORKS BY ROBERTSON

- (1915) 1948 *A Study of Industrial Fluctuation*. London: School of Economics and Political Science, Series of Reprints of Scarce Works on Political Economy, No. 8. London: Aldwych.
- (1922) 1959 *Money*. Rev. ed. Univ. of Chicago Press.
- (1923) 1960 ROBERTSON, DENNIS HOLME; and DENNISON, STANLEY R. *The Control of Industry*. New ed. Cambridge Economic Handbooks, No. 4. London: Nisbet. → Robertson was the sole author until the 1960 edition.
- (1924-1940) 1956 *Essays in Monetary Theory*. London: Staples.
- (1926) 1949 *Banking Policy and the Price Level*. Rev. ed. New York: Kelley.
- 1931 *Economic Fragments*. London: King.
- 1931 PIGOU, A. C.; and ROBERTSON, DENNIS HOLME *Economic Essays and Addresses*. London: King.
- 1952 *Utility and All That*. London: Allen & Unwin.
- 1954 *Britain in the World Economy*. London: Allen & Unwin.
- 1956 *Economic Commentaries*. London: Staples.
- 1957-1959 *Lectures on Economic Principles*. 3 vols. London: Staples. → A paperback edition was published in 1963 by Collins.
- 1961 *Growth, Wages, Money*. Cambridge Univ. Press.
- 1963 *Memorandum Submitted to the Canadian Royal Commission on Banking and Finance*. Essays in International Finance, No. 42. Princeton Univ., Department of Economics, International Finance Section.

## SUPPLEMENTARY BIBLIOGRAPHY

- ASHTON, T. S. 1951 *Industrial Fluctuation*. *Economica* New Series 18:298-302.
- FELLNER, WILLIAM J. 1952 *The Robertsonian Evolution*. *American Economic Review* 42:265-282.
- HICKS, J. R. 1965 Dennis Holme Robertson: 1890-1963. *British Academy, London, Proceedings* 50:305-316.
- MARGET, A. W. 1941 D. H. Robertson: *Essays in Monetary Theory*. *Review of Economic Statistics* 23:147-151.
- SAMUELSON, P. A. 1963 D. H. Robertson (1890-1963). *Quarterly Journal of Economics* 77:517-536.
- WILSON, T. 1953 Professor Robertson on Effective Demand and the Trade Cycle. *Economic Journal* 63:553-578.

## ROBERTSON SMITH, WILLIAM

See SMITH, WILLIAM ROBERTSON.

James Harvey Robinson (1863-1936), by common consent, did more than any other American historian to develop both academic and popular interest in the historiographical movement called the "New History." The New History had two principal and related characteristics. It repudiated the study of history for its own sake, placing it instead at the service of comprehension of the present and improvement of the future. Rather than concentrating on political and diplomatic events, the New History gave primary attention to social, economic, and intellectual developments.

**Education and influences.** Born of a well-to-do family in Bloomington, Illinois, Robinson had a rather unusual precollege education, his main interests at that stage apparently having been in science, notably biology and astronomy. He entered Harvard in 1883 and received his bachelor's degree in 1887, remaining a year for his master's degree. During this time he started work on what was to be his doctoral thesis, "The Original and Derived Features of the Constitution of the United States." He developed no absorbing enthusiasm for historical studies at Harvard, later stating that the only important intellectual inspiration he received there came from his work with William James in psychology and philosophy. Robinson studied for his doctorate in Germany under Hermann Eduard von Holst at the University of Freiburg. After he finished his doctoral work on the American constitution in 1890, von Holst interested him in German constitutional history, and in 1891 Robinson produced his *German Bundesrath*. These subjects were far removed from his later interests in general European intellectual and social history, but at this period of his life Robinson was really absorbed in constitutional history. However, the main permanent influence coming from his studies in Germany was his mastery of historical research and of the auxiliary sciences on which it depends: epigraphy, paleography, lexicography, and diplomatics.

Few historians have been as much affected by natural science as was Robinson, and it had a strong impact on his attitude toward history. His childhood interest in biology was continued and stimulated by his contacts with E. G. Conklin at the University of Pennsylvania (where he held his first teaching position) and with E. B. Wilson at Columbia (where Robinson taught later), as well as by summers at the marine biology laboratory at Woods Hole, Massachusetts. Robinson declared that he derived his genetic view of historical processes from biology rather than from historiography.



He demanded that history go beyond Leopold von Ranke's injunction to describe *what* happened and that it explain *how* and *why* it happened. Robinson was much impressed with the sensational advances that had recently been made in astronomy, and he believed that the new cosmic perspective had implications far more devastating to the traditional supernaturalism than even the teachings of biological evolution.

Those principally responsible for impressing on Robinson the importance of social and economic factors in historical development were Charles A. Beard, James T. Shotwell, Thorstein Veblen, and R. H. Tawney; Shotwell and Veblen, particularly, were inclined to emphasize the effects of technology and mechanization.

Related to Robinson's special concern with intellectual history was his interest in psychology, which dated from the strong impression that William James had made on him at Harvard. He gained a much more thorough, up-to-date, and technical knowledge of the field as the result of his long and close friendship with E. L. Thorndike at Columbia. In his later years Robinson became much interested in psychoanalysis, especially through his friendship with one of the leading analysts of that time, L. Pierce Clark, author of psychoanalytic studies of Napoleon and Lincoln. During their long friendship John Dewey fueled Robinson's interest in philosophical reflection on the course and meaning of history. Next to his interest in psychology the most potent factor encouraging Robinson's concern with intellectual history was his long, intensive, and appreciative study of the French rationalists.

The first person to have an important and enduring influence on Robinson's conception of history was Simon N. Patten, a welfare economist with a marked interest in the history of institutions, with whom Robinson worked at the University of Pennsylvania. Patten helped convince Robinson that the main value of history is its presentation of the past in a manner that facilitates understanding of the present and thereby permits a more rational approach to improvement of the future. Beard's early socialist inclinations also exerted some influence on Robinson's concern with the meliorative effect of historical knowledge. During his later years at Columbia Dewey's pragmatic philosophy, his leadership in progressive education, and his other strong reformist interests constituted the main influence on this aspect of Robinson's thinking. Robinson was never significantly affected by leaders in the New History abroad, such as Karl

Lamprecht and his disciples in Germany, Henri Berr, Charles Seignobos, and Alfred Rambaud in France, or Guglielmo Ferrero and Corrado Barbagallo in Italy.

**Career and traits as a teacher.** Robinson's professional career as a teacher of history was spent at the University of Pennsylvania, Columbia University, and, for a brief terminal period, at the New School for Social Research in New York City. While in Germany Robinson had met Patten and evidently made a very favorable impression on him, for Patten induced his university to invite Robinson as a lecturer in history. He was made an associate professor in 1892 and held that rank until he was called to Columbia in 1895 as a professor of European history.

The work at Pennsylvania for which he was best known grew out of his conviction that a command of source material is mandatory for any historian, for the practitioner of the New History as much as for the most haughty traditionalist in the profession. From the beginning of his teaching career, therefore, he believed strongly that history students should read source material, and to this end he edited several volumes of "Translations and Reprints from the Original Sources of European History" (see works edited by Robinson in bibliography).

Robinson's call to Columbia was the result of the determination of Dean John W. Burgess of the faculty of political science to add a specialist in European history. Robinson was selected mainly because William A. Dunning had been attracted by Robinson's emphasis on source material and by his skill in editing the "Translations and Reprints." Dunning also believed that Robinson could help him in his work as editor of the *Political Science Quarterly*, which proved to be true.

Aside from the final period in Robinson's instruction, during which he roamed fruitfully over the whole historical record of mankind in his presentation of intellectual history, his courses dealt with the Middle Ages, the Renaissance, the Reformation, and the French Revolution. His original and very influential interpretations of each of these periods or movements rank next to his promotion of the New History in determining his role in American historiography; indeed, they were an early phase of his New History and were embodied in his later lectures on intellectual history.

Robinson vigorously criticized the traditional view of the Middle Ages as a unique millennium of uniform gloom and cultural collapse following a mythical fall of Rome in 476, and he maintained

that there was probably no sharper break in the whole history of western Europe than that which separated the earlier from the later Middle Ages. He viewed the Renaissance as a slow and gradual development from late medieval advances and held that the conventional conception of it as a sudden flowering of a new culture was a myth. He taught that it is equally false to believe that the Reformation grew out of the Renaissance; instead, he emphasized the sharply contrasting traits of the two events and maintained that the Reformation was more a national political movement than a sudden religious uprising. He contended that the most fruitful approach to the French Revolution lay in a study of its causes and the course of its development, with less attention given to its dramatic and sensational episodes, which were transient and of little institutional import.

Robinson's method of presenting historical material to his classes varied notably during the course of his teaching career. At the outset he was very energetic and systematic and presented his material in a carefully organized way. Cheyney said that it "was his habit [to write] an outline of his lecture on the blackboard before he began to talk to the students. I know it made a great impression on them to feel that he had in mind exactly the group of things he wanted to say, classified in the way he wanted to say them" (Letter to Barnes, March 3, 1926). At Columbia he appears to have become gradually more relaxed and informal. According to Carl Becker, the famed Cornell historian, who was a student of Robinson's in his first decade at Columbia, "The professor talked so informally and entertainingly that taking notes seemed out of place. He had wit, a dry, mordant humor, and a fund of striking, unacademic bits of information which I had not found in textbooks or formal histories; and there was a sadness in the countenance, a quality, half plaintiveness half resignation, in the voice that made even simple statements of fact amusing or illuminating, or both" (Becker 1937, p. 48). Toward the end of his second decade at Columbia, Robinson's classroom popularity was perhaps unsurpassed in the teaching of history in the United States, despite a lecture method that would have made any other teacher appear very dull. Arthur M. Schlesinger, a student of Robinson's around 1910, has given us one of the best characterizations of Robinson's classroom manner at this period: "His lectures . . . proved the most provocative of any I attended and triggered endless arguments among the students. . . . Peering quizzically above his spectacles and impishly salt-

ing his observations with wit, he spoke quietly, as though interviewing himself, with the class present merely as eavesdroppers" (Schlesinger 1963, pp. 34-35).

In retrospect it appears most unfortunate that Robinson resigned from Columbia in 1919 rather than remaining there until he reached formal retirement age in 1928. His brief experience at the New School for Social Research was unsatisfactory; it did not give him opportunities to lecture to large classes and was not the type of educational experiment he had anticipated.

**The New History.** Since Robinson stands out in the evolution of historiography and historical writing primarily as the most widely known and effective exponent of the New History, his role in this area requires a rather precise examination.

His most readily visible accomplishment was the volume entitled *The New History* (1912), a book that exerted a wide and enduring influence on American historical writing and thinking. This influence was due partly to the attention it attracted by its appropriate title, partly to its convenient presentation of most of Robinson's observations on what the New History should be. But the contents of the book were not "new" even to Robinson. The first chapter, on the contrasts between the new—or cultural—history and the old—or political, episodic, and biographical—history, had first appeared as an article in 1900, and the succeeding chapters were articles and addresses published in the first decade of the twentieth century.

More than that, as Robinson himself was sufficiently well read in the history of historical writing to know, the New History was not his exclusive progeny. Although he was little affected by his European predecessors and contemporaries who exemplified the New History, he was well aware of such books as Andrew D. White's *A History of the Warfare of Science With Theology in Christendom*, Moses Colt Tyler's *Literary History of the American Revolution: 1763-1783*, H. C. Lea's work on the medieval church, Henry Adams' *Mont-Saint-Michel and Chartres*, and the work done by Frederick Jackson Turner and his students on the effect of the western frontier upon American culture and thought, to mention only a few of the authors who had already abandoned the older political and episodic history. What Robinson did was to christen the new departure in historical thought and writing, to found a movement around it, and actually to write exemplary history in terms of his conception of the subject.

Robinson's own development as the leading pro-



tagonist of the New History was a gradual one. As early as 1892, when he served on the Committee on Secondary School Studies of the National Education Association, he had enunciated what was to be a main tenet of his conception of the New History: that history should be written and taught in such a manner as to enable students to understand the conditions of their own time. He amplified this while serving on the 1904 Committee of the Association of History Teachers of the Middle States and Maryland, and he also advocated giving greater attention to modern history and world history. His collaboration with Beard on *The Development of Modern Europe* (Robinson & Beard 1907-1908) gave him a better comprehension of the importance of economic factors, especially of the industrial revolution, in the historical experience of mankind (see chapter 5 of *The New History* [1912] 1958, pp. 132-153).

Robinson's conception of the New History inevitably inclined him to regard history as a social science rather than as a branch of literature or an exercise in "collective biography." He was one of the first historians to call attention to the indispensable services of the social sciences to history, notably in his address before the American Historical Association in December 1910, entitled "The Relation of History to the Newer Sciences of Man," which constitutes the third chapter of *The New History* (*ibid.*, pp. 70-100). The fourth and the final chapters of *The New History* indicated Robinson's shift of interest to intellectual history, where he believed particularly strategic material could be found for the use of history in the service of social improvement. This was a theme he developed further about a decade later in two brief books, *The Mind in the Making* (1921), and *The Humanizing of Knowledge* (1923).

In his program of interpreting history and applying its lessons for the improvement of society, present and future, Robinson wrote many articles for historical periodicals in the 1920s and 1930s. These were brought together by Harry Elmer Barnes in a symposium entitled *The Human Comedy* (1937). The book is far and away the most comprehensive and revealing of all of the publications which present Robinson's contributions, containing as it does many of the important ideas that may be found in his other works and specific coverage of most of the leading topics and problems of modern history and contemporary life.

Textbooks. While it is rather commonly believed that Robinson's main contributions to the concepts and development of the New History are to be found in his writings about history, it seems

highly probable that he did far more to communicate his ideas about the nature and purpose of history by making them the core of his popular and widely used textbooks. The first and most revolutionary of these was *An Introduction to the History of Western Europe* (1902-1903), which was immediately followed by two volumes of *Readings in European History* (Robinson & Beard 1908-1909). Into the 1903 volume Robinson put his best thought and put it very well, this being the time of his maximum intellectual vigor and enthusiasm. This work set an entirely new standard for textbook writing in the historical field.

The outstanding aspect of the *Introduction to the History of Western Europe* lay in the author's successful discerning of the leading threads of medieval and modern European history and his tracing them with clarity and consistency. While he included all the political history that was necessary to understand the changing structures of European institutional life, he abandoned the traditional view that "history is past politics" and introduced ample material on social, cultural, and intellectual life. All this material was presented with Robinson's human touch and sprightly style, which gave vividness to the narrative and made it enjoyable to read. His readers were left unaware that Robinson's fluid prose was based on a rigorous study of the original sources which "controlled with relentless precision his reconstruction of the past," as his famed student, James Thomson Shotwell, insisted (Letter to Barnes, April 15, 1926).

Although the work ran counter to all precedent in the textbook field, it produced such a striking and favorable reaction on the part of history teachers that for more than a generation it dominated the college market. Over a quarter of a million copies of the first edition were sold. Admittedly, it was weak in treating the economic factors in history, especially in the modern period, and English constitutional history was given scanty attention. These defects were remedied in later editions after Robinson came into greater contact with Beard and collaborated with him in producing *The Development of Modern Europe*.

Robinson's influence in making the New History the basis of history teaching was extended to the secondary schools for many years through the two volumes of *Outlines of European History*, published in 1912-1914 (Robinson et al. 1912-1914). Robinson's collaborators in this instance were James H. Breasted of the University of Chicago and again Beard, with Breasted covering the ancient period and Beard giving primary attention to the treatment of economic factors in modern times. Revised

and adapted with the assistance of Emma Peters Smith and others, the work sold well until after Robinson's death. But there was little vital material in it that had not appeared in his college books, and by the 1920s, writers like Hutton Webster were producing texts that surpassed Robinson's in presenting the New History in fresh and original ways for the secondary schools.

**From teacher to preacher of history.** Between 1900 and 1912 Robinson was active and energetic in writing substantial, pioneer textbooks and in producing the articles and addresses that came to make up *The New History*. In the quarter of a century that followed 1912, an evident change took place, affecting both his professional life and his zest for writing. Robinson became wealthy and independent through his textbook royalties, but also increasingly relaxed in most matters, except for his ideas about history. Moreover, he loved to talk, was a witty and intriguing conversationalist, and spent more and more time talking to his favorite colleagues like Dewey, Thorndike, and Wilson. Beard was inclined to think that this did more than anything else to distract Robinson from further substantial historical publications, and may have accounted for his failure to produce his expected magnum opus on the history of the intellectual class in western Europe. Although Robinson was still a relatively young man in 1907 when *The Development of Modern Europe* (Robinson & Beard 1907-1908) was published, he never produced another historical volume of the scale and quality of this book and of his original *Introduction to the History of Western Europe*.

As Robinson devoted himself more and more to thinking and talking about history rather than writing it, he gave increased attention to interpreting history and to considering what service it might render to the understanding and betterment of mankind. As Lynn Thorndike put it,

The chief change that came over Professor Robinson in the course of his life . . . was that from being a teacher of history he became a preacher of history. The small seminars and informal classes gave way to crowded lectures and these to a host of readers. . . . He would humanize knowledge. While some of us are content with the hope that our ideas and findings may gradually filter down through our students and learned publications to have their ultimate effect on human thought and society, Robinson was impelled to make more direct contacts. Despite his critical attitude toward past and existing religion, he had the faith not merely that would move mountains but would move muttonheads. (Thorndike 1937, p. 369)

**The Columbia school of New Historians.** Robinson's associates were tied together not so much by

their common interest in a particular subject of investigation but by a common attitude toward the nature and purpose of history in general and by a sort of filial devotion to Robinson. The nature of the so-called "Robinson school" has been well stated by Crane Brinton: ". . . around Mr. Robinson and other leaders a definite, self-conscious school of historians has formed, and . . . this school has cohesion, common purposes, common watchwords, common loves and hatreds . . . in short, it is a school in the good, loose sense in which the Stoa, the Pleiad, the Lake poets, or the Utilitarians were schools" (Brinton 1936, p. 134).

Writers from this school produced creditable, and in some cases impressive, works in many fields. In the area of intellectual history probably the most important were the able editorial work and translations of church and medieval historical works and documents by Louise R. Loomis, Ernest Brehaut, and Irving W. Raymond; the massively erudite volumes of Lynn Thorndike on the history of European magic and science; Alexander C. Flick's substantial treatment of the medieval church; the definitive writings of Preserved Smith on the Renaissance and Reformation and the intellectual life of the seventeenth and eighteenth centuries; Dorothy Stimson's account of the impact of the Copernican system; Martha Ornstein's brilliant work on the rise of the scientific societies; Becker's influential interpretations of the intellectual backgrounds of the English, American, and French revolutions; the work of Howard Robinson on French rationalism; Anne E. Burlingame's study of Condorcet; J. Salwyn Schapiro's treatment of modern French liberal and revolutionary thought; Carlton Hayes's writings on the nature, ramifications, and influence of nationalism; and J. H. Randall, Jr.'s monumental history of European philosophy and thought. The Robinsonian emphasis on social history was applied in American as well as in European history, by Beard, Schlesinger, Harold U. Faulkner, Dixon Ryan Fox, and Harry J. Carman. Robinson's interests in historiography were carried on mainly by Shotwell and Barnes, with some incidental but valuable contributions by Schlesinger and Henry Johnson. The only one of Robinson's disciples who gave special attention to social reform and amelioration was Barnes, who did this as much in his role as sociologist and criminologist as in that of historian.

It was Barnes also who produced the only comprehensive synthetic works in the Robinsonian tradition, of which the first was *A History of Western Civilization* (Barnes & David 1935). Robinson himself lived to read, review, and approve of this



work. Crane Brinton, although somewhat critical of the Robinsonians, characterized the book as "this advertised inheritance, this harvest of a long tradition, this long-awaited fruition" of the Columbia school (Brinton 1936, p. 135); and Preserved Smith described it as "incontestably the masterpiece of the New History" (Letter to Barnes, Sept. 6, 1935). Barnes next endeavored to fill the gap created by Robinson's failure to produce his intellectual history. Aided by experts in the history of art, literature, and music—subjects with which Robinson did not deal extensively in his writings—Barnes produced the *Intellectual and Cultural History of the Western World* (1937), a more comprehensive work than Robinson had ever planned.

In 1929 a symposium entitled *Essays in Intellectual History*, containing contributions by some of Robinson's leading students in this field, was presented to him at the time of his election to the presidency of the American Historical Association. This testimonial volume attested to the distinction of his students and to the diversity of interests that he stimulated.

As an appraisal of the significance of the Robinsonian epoch in historical writing one may conclude with the critical and restrained summation by Brinton:

Now the battle to free history from the limitation of the doctrine that history is "past politics" . . . has been won, and won largely through the efforts of the Columbians. With all their faults, the history text-books of today afford a far more nourishing fare than did the text-books which Mr. Robinson ridiculed so effectively in *The New History*. . . . The battle against purely political history was worth fighting, and the victory worth winning. (Brinton 1936, p. 135)

HARRY ELMER BARNES

[Directly related are the entries HISTORY, articles on SOCIAL HISTORY and INTELLECTUAL HISTORY; and the biographies of BARNES; BEARD; DEWEY; JAMES; RANKE; THORNDIKE; VEBLEN.]

#### WORKS BY ROBINSON

- 1890 *The Original and Derived Features of the Constitution*. American Academy of Political and Social Science, *Annals* 1:203-243.
- 1891 *The German Bundesrath: A Study in Comparative Constitutional Law*. University of Pennsylvania Political Economy and Public Law Series, Vol. 3, No. 1. Philadelphia: Univ. of Pennsylvania Press.
- (1902-1903) 1946 *An Introduction to the History of Western Europe*. Rev. & enl. by James T. Shotwell. 2 vols. Boston: Ginn.
- 1904-1906 *Readings in European History: A Collection of Extracts From the Sources Chosen With the Purpose of Illustrating the Progress of Culture in Western Europe Since the German Invasions*. 2 vols. Boston: Ginn. → Volume 1: *From the Breaking Up of the*

*Roman Empire to the Protestant Revolt*. Volume 2: *From the Opening of the Protestant Revolt to the Present Day*.

- (1907-1908) 1929-1930 ROBINSON, JAMES HARVEY; and BEARD, CHARLES A. *The Development of Modern Europe: An Introduction to the Study of Current History*. Rev. & enl. ed. Boston: Ginn.
- 1908-1909 ROBINSON, JAMES HARVEY; and BEARD, CHARLES A. *Readings in Modern European History: A Collection of Extracts From the Sources Chosen With the Purpose of Illustrating Some of the Chief Phases of the Development of Europe During the Last Two Hundred Years*. 2 vols. Boston: Ginn. → Volume 1: *The Eighteenth Century: The French Revolution and the Napoleonic Period*. Volume 2: *Europe Since the Congress of Vienna*
- (1911) 1919 *An Outline of the History of the Western European Mind*. 4th ed., rev. New York: Marion. → First published as *An Outline of the History of the Intellectual Classes in Western Europe*
- (1912) 1958 *The New History: Essays Illustrating the Modern Historical Outlook*. Springfield, Mass.: Walden. → See especially pages 70-100 on "The New Allies in History" and pages 132-153 on "History for the Common Man."
- (1912-1914) 1914-1918 ROBINSON, JAMES HARVEY et al. *Outlines of European History*. 2 vols. Boston: Ginn.
- 1918 *The Last Decade of European History and the Great War*. Boston: Ginn. → Designed as a supplement to *The Development of Modern Europe: An Introduction to the Study of Current History* by James Harvey Robinson and Charles A. Beard and to *An Introduction to the History of Western Europe* by James Harvey Robinson.
- (1921) 1950 *The Mind in the Making: The Relation of Intelligence to Social Reform*. With an introduction by Stuart Chase. New York: Harper.
- (1921) 1934 ROBINSON, JAMES HARVEY; and BEARD, CHARLES A. *History of Europe, Our Own Times, the Eighteenth and Nineteenth Centuries: The Opening of the Twentieth Century, the World War and Recent Events*. Rev. ed. Boston: Ginn.
- (1921) 1934 ROBINSON, JAMES HARVEY; and SMITH, EMMA P. *Our World Today and Yesterday: A History of Modern Civilization*. Boston: Ginn. → First published as *A General History of Europe From the Origins of Civilization to the Present Time*.
- (1923) 1926 *The Humanizing of Knowledge*. 2d ed., rev. New York: Doran.
- 1926 *The Ordeal of Civilization: A Sketch of the Development and World-wide Diffusion of Our Present-day Institutions and Ideas*. New York: Harper.
- 1937 *The Human Comedy as Devised and Directed by Mankind Itself*. With an introduction by Harry Elmer Barnes. New York: Harper. → Published posthumously.

#### WORKS EDITED BY ROBINSON

- (1894) 1897 ROBINSON, JAMES HARVEY (editor) *The French Revolution: 1789-1791*. Rev. ed. Translations and Reprints, Vol. 1, No. 5. Philadelphia: Univ. of Pennsylvania, Department of History.
- 1895 ROBINSON, JAMES HARVEY (editor) *The Napoleonic Period*. Translations and Reprints, Vol. 2, No. 2. Philadelphia: Univ. of Pennsylvania, Department of History.
- 1895 ROBINSON, JAMES HARVEY; and WHITCOMB, MERICK (editors) *The Period of the Early Reformation in Germany*. Translations and Reprints, Vol. 2, No. 6.

Philadelphia: Univ. of Pennsylvania, Department of History.

- 1896 ROBINSON, JAMES HARVEY (editor) *The Restoration and the European Policy of Metternich: 1814-1820*. Translations and Reprints, Vol. 1, No. 3. Philadelphia: Univ. of Pennsylvania, Department of History.
- 1897 ROBINSON, JAMES HARVEY (editor) *The Pre-reformation Period*. Translations and Reprints, Vol. 3, No. 6. Philadelphia: Univ. of Pennsylvania, Department of History.
- 1899 ROBINSON, JAMES HARVEY (editor) France, Cour des aides, *Protest of the Cour des aides of Paris, 10 April 1775*. Translations and Reprints, Vol. 5, No. 2. Philadelphia: Univ. of Pennsylvania, Department of History.

#### SUPPLEMENTARY BIBLIOGRAPHY

- [Book Review of] *An Introduction to the History of Western Europe*, by James Harvey Robinson. 1903 *Nation* 76:502-503.
- BARNES, HARRY E. (1927) 1965 James Harvey Robinson. Pages 319-408 in Howard W. Odum (editor), *American Masters of Social Science: An Approach to the Study of the Social Sciences Through a Neglected Field of Biography*. Port Washington, N.Y.: Kennikut.
- BARNES, HARRY E. (1937) 1961 *Intellectual and Cultural History of the Western World*. 3d rev. ed., 3 vols. New York: Reynal-Hitchcock. → A paperback edition was published by Dover in 1965.
- BARNES, HARRY E.; and DAVID, HENRY 1935 *A History of Western Civilization*. 2 vols. New York: Harcourt.
- BECKER, CARL 1937 James Harvey Robinson. *Nation* 144:48-50.
- BRINTON, CRANE 1936 The New History: Twenty-five Years After. *Journal of Social Philosophy* 1:134-147.
- Essays in Intellectual History: Dedicated to James Harvey Robinson by His Former Seminar Students*. 1929 New York: Harper.
- HENDRICKS, LUTHER V. 1946 *James Harvey Robinson: Teacher of History*. New York: King's Crown Press.
- JOHNSON, HENRY 1943 *The Other Side of Main Street: A History Teacher From Sauk Centre*. New York: Columbia Univ. Press.
- MILLER, ALICE (Duer); and MYERS, SUSAN 1939 *Barnard College: The First Fifty Years*. New York: Columbia Univ. Press.
- SCHAPIRO, J. SALWYN 1936 James Harvey Robinson (1863-1936). *Journal of Social Philosophy* 1:278-281.
- SCHLESINGER, ARTHUR M. 1963 *In Retrospect: The History of a Historian*. New York: Harcourt.
- SELLERY, GEORGE S. 1904 [Book Review of] *An Introduction to the History of Western Europe*, by James Harvey Robinson. *American Academy of Political and Social Science, Annals* 23:165-167.
- THORNDIKE, LYNN 1937 [Book Review of] *The Human Comedy as Devised and Directed by Mankind Itself*, by James Harvey Robinson. *Journal of Modern History* 9:367-369.

#### RODBERTUS, JOHANN KARL

Johann Karl Rodbertus (1805-1875), also known as Rodbertus-Jagetzow, was born in Greifswald, Pomerania, then under Swedish jurisdiction, where his father was a professor of Roman law and a member of the Swedish judiciary. His

mother was the daughter of Johann August Schlettwein, a physiocrat and professor of statistics and political economy. After Napoleon's troops occupied Greifswald, the elder Rodbertus resigned his post, in 1808, and withdrew with his family to one of the Schlettwein estates. Rodbertus then attended the Gymnasium in Friedland, Mecklenburg, an institution permeated with the liberalism of that period. His subsequent study of law in Göttingen and Berlin, from 1823 to 1826, led to his service as a barrister to the Prussian state. In 1832 he left government service and turned to the study of political economy, first in Dresden and then in Heidelberg, supplementing his academic studies with travel in Switzerland, France, and Holland. There he became aware of the grave social tensions related to industrialization, and these remained his concern to the end of his life.

Rodbertus was a German state socialist, as were Ferdinand Lassalle and Adolf Wagner. He was in constant correspondence with these two men, and their mutual influence is unmistakable, particularly that between Rodbertus and Lassalle. These two agreed that given the existing political and legal order and the institution of private property in capital and land that was based on it, there was no hope of solving the "social problem" (i.e., the social disorganization attendant on the industrial revolution). Yet Rodbertus and Lassalle differed in their approaches to the modification of the existing order: Lassalle sought to achieve his aims by forming a political party—a social democratic party—while Rodbertus insisted that the socialist movement concern itself only with economic matters and abstain from all political aims.

Rodbertus' diagnosis of the ills of society begins with an analysis of the economic and social effects of private property. In an economic system based on private property, labor becomes a commodity and land becomes capital. If the private ownership of land and the means of production were eliminated, all goods might, from the economic point of view, be considered products of labor. Their value, then, would depend exclusively on the amount of labor required for their production. The existence in the economic system of capital rent (interest) and land rent was attributed by Rodbertus to two causes, one economic and one legal. The economic cause is the fact that the division of labor and advancing production technology enable the worker to produce more than he requires for his own support. The legal cause lies in the protection by law of private ownership of land and of the means of production. Both the peasants and the workers may therefore be



charged by the owners for the use of their property, and the charge that is exacted takes the form of part of the product of labor. Rodbertus called this "rental profit," which in turn can be divided into land rent and capital rent.

According to Rodbertus' theory of labor value, capital rent is independent of the total capital, deriving only from that part of capital which is expended on labor. Since rent represents the ratio of the surplus value levied to the capital invested by the entrepreneur, it is lower in industry than in agricultural production, where the soil, as a gratuitous means of production, contributes its own fertility to output. Hence, what Rodbertus called an "absolute land rent" is derived from the soil, and the soil should be considered as a rent fund rather than as capital. The implication of this theory of rent for agrarian policy is, according to Rodbertus, that land should not be encumbered with capital debt but only with the obligation of paying rent. He believed that mortgage indebtedness should be replaced by payment of permanently fixed rent. This proposal was effected in part, and only temporarily, by the establishment of so-called rent estates.

As Rodbertus saw it, the moral development of man is impeded by a social system that, by virtue of private ownership of land and of the means of production, turns labor into a commodity and land into capital. For the sake of the community as a whole, over and above that of the individual, it was necessary to change the system. Unlike Marx, Rodbertus did not believe that change had to be revolutionary; his conception of change was influenced by Fichte and Schelling and was evolutionary in nature. To be sure, Rodbertus did not believe that mankind would simply evolve toward an ideal type; rather, he thought that ever-increasing economic entanglements would impel men to effect remedial political action. Only in the remote future would the process of historical evolution result in a society organized on a moral basis, with the total elimination of private ownership of the means of production and of land. Such a social-historical conception—it has also been termed metaphysical collectivism—was hardly practical, and it is not surprising that Rodbertus' doctrines had little or no influence on the German labor movement.

The basic principles of Rodbertus' social philosophy (later supplemented by his political economy) were first presented in an article entitled "Die Forderungen der arbeitenden Klassen" (submitted to the *Augsburger Allgemeine Zeitung* in 1837 or 1839, but not accepted by that news-

paper). Rodbertus believed that a social order must be based on one of two systems: either that of discipline and subordination, or that of education (which in turn is based on upbringing and instruction). In antiquity, the social order was of necessity based on discipline and subordination; Rodbertus believed slavery and exploitation to be justified as long as the productivity of the economy was so low as to support adequately only the cultured stratum of society. But with the growth of productivity, the poorer strata should be granted a share in the social product over and above the subsistence level. If this does not occur and only the few continue to benefit from the increased flow of goods, it is because obsolete juridical and economic institutions are vitiating proper distribution, and it means that the social order requires modification.

What makes for this "wrong" distribution is the iron law of wages, by virtue of which the social product is so distributed among landowners, capitalists, and workers that the last receive a constant amount, and hence a constantly diminishing share, despite rising productivity. In an age when discipline and slavery no longer serve as the underlying principles of social organization but have been replaced by education, nothing can save society from collapse so long as the workers' misery and poverty "destroy what the school is trying to accomplish" (1839, p. 16). There is then a fundamental defect in those legal and economic institutions that permit the iron law of wages to drive the classes of society increasingly apart and thus to imperil the progress of civilization and the very existence of culture.

Only government intervention, Rodbertus believed, can counteract what he called the law of the decreasing wage share. The rent received by capitalists and landowners should be abolished, thus eliminating "rent-receiving private property," at the same time that the workers' ownership of the product of labor is guaranteed. Rodbertus proposed that the average labor time spent on production determine the distribution of the social product. (For Rodbertus' related concept of the "normal working day," see his 1871 article.) Each worker should receive a confirmed claim to products collected in a storehouse, corresponding to his contribution in terms of labor time. A new currency of constant value can be introduced in this manner. (It is worth noting the failure of Robert Owen's attempt, in 1833, to establish a labor exchange.)

Rodbertus' researches in economic history served to support various aspects of his theoretical system—for example, his assertion that originally

there was only a single rent, which became differentiated into land rent and interest on capital with the introduction of the division of labor. His conception of the historical relativity of economic conditions led him to reject revolutionary demands, on theoretical as well as political grounds. Politically, he considered every revolution a detour, since it involved the destruction of law and order; as a theorist, he advocated a guided "system of social foresight" in lieu of *laissez-faire*. He was not in favor of tampering with the current distribution of property; only its "fruits" (interest on capital and land) should be redistributed by governmental measures.

BRUNO FRITSCH

[See also RENT; and the biographies of THÜNEN; WAGNER.]

#### WORKS BY RODBERTUS

- (1839) 1946 *Die Forderungen der arbeitenden Klassen*. Sozialökonomische Texte, Vol. 5. Frankfurt am Main: Klostermann.
- 1842 *Zur Erkenntnis unserer staatswirtschaftlichen Zustände*. Neubrandenburg: Lingnau.
- (1850) 1898 *Overproduction and Crises*. New York: Scribner. → First published in German as the second of the "Soziale Briefe an von Kirchmann."
- 1864-1867 Untersuchungen auf dem Gebiete der Nationalökonomie des klassischen Alterthums. Parts 1-2. *Jahrbücher für Nationalökonomie und Statistik* 2:206-268; 4:341-427; 5:135-171, 241-315; 8:81-126, 385-475. → Part 1: Zur Geschichte der agrarischen Entwicklung Roms unter den Kaisern, oder die Adscriptitii, Inquilinen und Colonen. Part 2: Zur Geschichte der römischen Tributsteuern seit Augustus.
- (1868-1869) 1893 *Zur Erklärung und Abhülfe der heutigen Creditnoth des Grundbesitzes*. 2d ed. Berlin: Bahr.
- 1870 *Zur Frage des Sachwerthes des Geldes im Alterthum*. *Jahrbücher für Nationalökonomie und Statistik* 14: 341-420; 15:182-234.
- (1871) 1899 *Der Normal-Arbeitstag*. Pages 337-359 in Johann Karl Rodbertus, *Schriften*. Volume 4: Gesammelte kleine Schriften. Berlin: Puttkammer & Mühlbrecht.
- 1884 Ein Versuch, die Höhe des antiken Zinsfusses zu erklären. *Jahrbücher für Nationalökonomie und Statistik* 42:513-535. → Published posthumously.
- 1891 *Erinnerungen an Rodbertus*. Part 1: Eine autobiographische Skizze. *Jahrbuch für Gesetzgebung, Verwaltung und Volkswirtschaft im Deutschen Reich* 15: 585-591. → Published posthumously.
- Aus dem literarischen Nachlass von Carl Rodbertus-Jagetzow. Volume 1: Briefe von Ferdinand Lassalle an Carl Rodbertus-Jagetzow. Berlin: Puttkammer & Mühlbrecht, 1878.
- Briefe und sozialpolitische Aufsätze*. 2 vols. Edited by Rudolph Meyer. Berlin: Klein, 1882.
- Deutscher Staat und Sozialismus*. Selected and with a foreword by Horst Wagenführ. Potsdam: Protte, 1935.
- Neue Briefe über Grundrente, Rentenprinzip und soziale Frage an Schumacher*. Edited by Robert Michels and Ernst Ackermann. Bibliothek der Soziologie und Politik, Vol. 1. Karlsruhe: Braun, 1926.
- Schriften von Johann Carl Rodbertus Jagetzow*. 4 vols. Edited by Adolph Wagner, Theophil Kozak, and Moritz Wirth. Berlin: Puttkammer & Mühlbrecht, 1899. → Volume 1: *Das Kapital* (1884). Volumes 2-3: *Zur Beleuchtung der sozialen Frage*. 2 vols. (1850-1885). Volume 4: *Gesammelte kleine Schriften* (1845-1871).

#### SUPPLEMENTARY BIBLIOGRAPHY

- ADLER, GEORG 1884 *Rodbertus: Der Begründer des wissenschaftlichen Sozialismus*. Leipzig: Duncker & Humblot.
- BAHR, HERMANN 1884 *Rodbertus' Theorie der Absatzkrisen*. Vienna: Konegen.
- BRADKE, ELISABETH VON 1923 *Die Gesellschaftslehre von Karl Rodbertus-Jagetzow*. *Archiv für Sozialwissenschaft und Sozialpolitik* 50:34-86.
- CONRAD, J. 1870 *Das Rentenprinzip nach Rodbertus' Vorschlag und seine Bedeutung für die Landwirtschaft*. *Jahrbücher für Nationalökonomie und Statistik* 14: 149-182.
- CORNÉLISSSEN, CHRISTIAN (1903) 1926 *Théorie de la valeur, avec une réfutation des théories de Rodbertus & Karl Marx, Walras, Stanley Jevons & Boehm-Bawerk*. 2d ed. Paris: Giard.
- DIETZEL, HEINRICH 1886 *Das "Problem" des literarischen Nachlasses von Rodbertus-Jagetzow*. *Jahrbücher für Nationalökonomie und Statistik* 47:243-257.
- DIETZEL, HEINRICH 1886-1888 *Karl Rodbertus: Darstellung seines Lebens und seiner Lehre*. 2 vols. Jena: Fischer.
- GRAZIANI, AUGUSTO 1927 *Le lettere inedite di Rodbertus*. *Giornale degli economisti e rivista di statistica* Fourth Series 67:195-198.
- HERBERG, KÄTHE 1929 *Die Einkommenslehre bei Thünen und bei Rodbertus*. Hagen: Ose.
- KOLISCHER, HEINRICH 1876 *Rodbertus' Ansichten über den landwirtschaftlichen Hypothekenkredit: Erstes Hauptstück einer Arbeit über Verschuldung und Ueber-schuldung des ländlichen Grundbesitzes*. Vienna: Seidl-Mayer.
- KOZAK, THEOPHIL 1879 *Verzeichniss der hauptsächlichsten Publikationen von Rodbertus*. *Jahrbücher für Nationalökonomie und Statistik* 33:311-324.
- KOZAK, THEOPHIL 1882 *Rodbertus-Jagetzow's sozialökonomische Ansichten*. Jena: Fischer.
- LEXIS, W. 1884 *Zur Kritik der Rodbertus'schen Theorien*. *Jahrbücher für Nationalökonomie und Statistik* 43 462-476.
- MICHELS, ROBERT 1926 *Rodbertus und sein Kreis*. Pages 1-82 in Karl Rodbertus, *Neue Briefe über Grundrente, Rentenprinzip und soziale Frage an Schumacher*. Karlsruhe: Braun.
- MUZIOL, ROMAN 1927 *Karl Rodbertus als Begründer der sozialrechtlichen Anschauungsweise*. Beiträge zur Geschichte der Nationalökonomie, Vol. 4. Jena: Fischer.
- NACHT, OSKAR 1907 *Rodbertus' Stellung zur sozialen Frage*. Berlin: Puttkammer & Mühlbrecht.
- OPPENHEIMER, FRANZ 1908 *Rodbertus' Angriff auf Ricardos Renten-Theorie und der Lexis-Diehl'sche Rettungsversuch*. Berlin: Reimer.
- RUEHKOPF, KARL 1892 *Rodbertus Theorie von den Handelskrisen; Darstellung und Kritik: Eine Studie*. Leipzig: Gräfe.
- SCHULTZ, WILHELM 1882 *Die Arbeit als Quelle und Maass des Werthes*. Leipzig: Findel.



- STORCH, LUDWIG 1930 *Karl Rodbertus und Adolph Wagner als Staatsozialisten*. Giessen: Engel.
- THIER, ERICH 1930 *Rodbertus, Lassalle, Adolph Wagner: Ein Beitrag zur Theorie und Geschichte des deutschen Staatsozialismus*. Jena: Fischer.
- WANSTRAT, RENATE 1950 *Johann Karl Rodbertus: Zur 75. Wiederkehr seines Todestages. Schmollers Jahrbuch für Gesetzgebung, Verwaltung und Volkswirtschaft* 70:385-408.
- WARNINGHOFF, ALBERT 1938 *Die Bestrebungen um die Neugestaltung des landwirtschaftlichen Kredits seit Rodbertus*. Hamburg: Preilipper.
- ZUNS, JULIUS 1883 *Etwas über Rodbertus*. Berlin: Puttkammer & Mühlbrecht.

## ROGERS, JAMES E. THOROLD

James Edwin Thorold Rogers (1823-1890), one of the founders of modern British economic history, was a vigorous and versatile scholar whose forceful personality and uninhibited political partisanship directly affected his academic career. As an interpretative historian, he consistently stressed the economic basis of political issues. His personality is evident throughout his historical writings, especially in his prefaces; he repeatedly denounced the follies and injustices of his own and previous generations and boldly asserted his claims to originality and pre-eminence in his chosen field of research.

Rogers' life began conventionally enough. He was born at West Meon, Hampshire, and after studying at Southampton School and King's College, London, he graduated from Magdalen College, Oxford in 1846, the year of Richard Cobden's free-trade triumph. An ardent High-churchman, he was ordained at Oxford, where he combined his clerical duties with a successful career as private tutor in classics and philosophy, examiner, and holder of various unpaid university administrative posts. About 1860, however, the whole pattern of his life changed. He abandoned the church because of his increasing dissatisfaction with the Tractarian movement and became a leading supporter of the Clerical Disabilities Relief Act of 1870—which he was the first to avail himself of. He turned from the classics to economic subjects, partly in response to suggestions made at the International Statistical Congress and partly because the Clarendon Press had refused to publish his Aristotelian dictionary. In 1861 he published *Education in Oxford: Its Methods, Its Aids, and Its Rewards*, which contained many severe criticisms of the university. He had already, in 1859, been appointed Tooke professor of statistics and economic science at King's College, London, a post he retained until his death, and in 1862 he also became Drummond professor

of political economy at Oxford, nominally a five-year appointment. Rogers' tenure was characteristically energetic. He delivered frequent lectures enlivened by racy anecdotes and incisive comments on contemporary personalities and issues, and he published the first two volumes of his monumental *History of Agriculture and Prices in England* in 1866, as well as other works on political economy, ethics, law, and politics. Nevertheless, he failed to secure re-election in 1868, owing to a vigorous campaign by members of convocation who objected to his strictures on Oxford and his outspoken radical opinions.

This defeat reinforced Rogers' iconoclasm and drew him inexorably toward politics, a field which brought out the boisterous and intemperate side of his nature. His political ideas were deeply influenced by his intimacy with Richard Cobden, whose sister had married Rogers' elder brother. Rogers edited Cobden's and Bright's speeches (see Cobden 1870; Bright 1879), and the essays in his *Cobden and Modern Political Opinion* (1873) reveal his staunch adherence to free trade. In 1880, after an earlier unsuccessful candidacy, he entered the House of Commons and remained a member for six years. But he was neither a conspicuously successful nor active parliamentarian, for he continued to live at Oxford and pursue his scholarly activities. His most important writings include *Six Centuries of Work and Wages* (1884), which contains the preliminary results of his indefatigable researches, and four additional volumes of his *History of Agriculture and Prices*, which appeared in 1887. Volume 7, in two parts, was published posthumously, as was a volume of lectures, *The Industrial and Commercial History of England* (1892). These, like his earlier lectures entitled *The Economic Interpretation of History* (1888a), were delivered at Oxford, where Rogers was re-elected to the Drummond chair in 1888, following the death of Bonamy Price, his rival of the 1860s. This reinstatement was widely interpreted as an effort to make restitution for the earlier injustice, a verdict which Rogers himself did nothing to discourage.

As a scholar Rogers was too wayward, too much influenced by personal likes and dislikes. His scathing comments on Ricardo's and Malthus' "metaphysical" school of English political economy and his insistence on the need to base economic principles on an exhaustive examination of the facts made him a natural ally of the historical economists, who were attacking the doctrines and methods of classical economics during the 1870s and 1880s. Yet he was far too idiosyncratic to be a

faithful member of any doctrinal school. Indeed, he greatly exaggerated his disagreements with Ricardo and Malthus, and in contrast to the historicists, he was profoundly influenced by Frédéric Bastiat's laissez-faire ideas. Admittedly, he qualified his earlier uncritical adherence to the wages-fund doctrine and became a warm advocate of trade unions; but he made only minor concessions to those who desired an increase in government intervention in economic and social life.

As a historian Rogers was a great pioneer, but he lacked the indispensable qualities of modesty and caution. In contrast to his painstaking accumulations of price data, which he published with indiscriminating zeal, he made many sweeping and uncritical judgments about the past. His conviction that "all genuine facts are far more valuable than the inferences of any individual who uses them" (1866–1902, vol. 4, p. vi) was, therefore, borne out in his own case. At a time when historians were being increasingly influenced by evolutionary ideas, Rogers' approach to the past was, as one contemporary critic observed, "cataclysmic" (Ashley 1889, p. 400; see also Gibbins 1891). He seriously overrated the pernicious influence of governmental policies, past and present, but his hostility toward all forms of privilege, his distaste for the governing classes, and his sympathy toward the workers made his books acceptable to several generations of students, especially in the world of adult education. Through this medium, in conjunction with Arnold Toynbee, Rogers helped to popularize the pessimistic view of the consequences of the industrial revolution that dominated English economic history almost until the 1940s.

A. W. COATS

[For the historical context of Rogers' work, see ECONOMIC THOUGHT, article on THE HISTORICAL SCHOOL; HISTORY, article on ECONOMIC HISTORY; and the biographies of BASTIAT and TOYNBEE.]

#### WORKS BY ROGERS

- 1861 *Education in Oxford: Its Method, Its Aids, and Its Rewards*. London: Smith & Elder.
- 1864 *Primogeniture and Entail: Letters of J. E. T. Rogers and Henry Tupper and Others, on the History and Working of the Laws of Primogeniture and Entail, in Their Moral, Social and Political Aspects*. Manchester (England): Ireland.
- 1866–1902 *A History of Agriculture and Prices in England, From the Year After the Oxford Parliament (1259) to the Commencement of the Continental War (1793)*. 7 vols. Oxford: Clarendon.
- 1867 Bribery. Pages 259–276 in Frank H. Hill et al., *Questions for a Reformed Parliament*. London: Macmillan.
- (1868) 1876 *A Manual of Political Economy for Schools and Colleges*. 3d ed. rev. Oxford: Clarendon.

- 1869–1870 *Historical Gleanings*. 2 vols. First and Second Series. London: Macmillan.
- 1873 *Cobden and Modern Political Opinion*. London: Macmillan.
- 1883 *Ensilage in America: Its Prospects in English Agriculture*. London: Sonnenschein.
- (1884) 1949 *Six Centuries of Work and Wages: The History of English Labour*. Edited with a new preface by G. D. H. Cole. London: Allen & Unwin.
- 1887 *The First Nine Years of the Bank of England: An Enquiry Into a Weekly Record of the Price of Bank Stock From August 17, 1694 to September 17, 1703*. New York: Macmillan.
- 1888a *The Economic Interpretation of History*. New York: Putnam.
- 1888b *The Relations of Economic Science to Social and Political Action*. London: Sonnenschein.
- (1888c) 1900 *Holland*. New York: Putnam.
- 1892 *The Industrial and Commercial History of England*. Edited by Arthur G. L. Rogers. New York: Putnam. → Published posthumously.

#### WORKS EDITED BY ROGERS

- ARISTOTLE *Aristotelis ethica nicomachea*. London: Rivington, 1865.
- BRIGHT, JOHN 1879 *Public Addresses*. London: Macmillan.
- CODDEN, RICHARD (1870) 1880 *Speeches on Questions of Public Policy*. Edited by John Bright and James E. Thorold Rogers. London: Macmillan.
- GREAT BRITAIN, PARLIAMENT, HOUSE OF LORDS 1875 *Complete Collection of the Protests of the Lords, With Historical Introductions*. Edited from the Journals of the Lords. 3 vols. Oxford: Clarendon.
- OXFORD (CITY) 1891 *Oxford City Documents: Financial and Judicial: 1268–1665*. Oxford Historical Society Publication, No. 18. Oxford: Clarendon.
- SMITH, ADAM (1776) 1869 *An Inquiry Into the Nature and Causes of the Wealth of Nations*. Oxford: Clarendon.

#### WORKS ABOUT ROGERS

- ASHLEY, W. J. 1889 James E. Thorold Rogers. *Political Science Quarterly* 4:381–407.
- GIBBINS, H. B. DE 1891 In Memoriam: Thorold Rogers. *Economic Review* (London) 1:86–89.
- HEWINS, W. A. S. (1897) 1921 James Edwin Thorold Rogers. Volume 17, pages 123–126 in *Dictionary of National Biography*. London: Oxford University Press.

## RÓHEIM, GÉZA

Géza Róheim, anthropologist and psychoanalyst, was born in Budapest in 1891 and died in New York in 1953. Róheim came from a family of wealthy merchants. Even in his preschool years he showed a more than ordinary interest in myths and fairy tales, and while still a high school student he became an expert in Hungarian folklore. At the age of 18 he wrote an essay, "Dragons and Dragon Killers" (1911), that presaged some of his later psychological and cultural theories. Róheim's interest in mythology and folklore continued throughout his life: when he was only five he read *Alice*



in *Wonderland* and the *Arabian Nights*, and in his last major work, *The Gates of the Dream* (1952), he discussed what he considered to be one of the basic themes of *Alice*, her fall into the rabbit hole.

Róheim attended the University of Budapest, until his interest in ethnographic research led him to the University of Leipzig to study under Karl Weule and to the University of Berlin to study under Felix von Luschan. While in Germany, he became acquainted with Freud's writings, including the psychoanalytic interpretation of culture first proposed in *Totem and Taboo*.

After having received his PH.D., Róheim returned to Budapest, where he became a staff member of the Hungarian National Museum. It was at this time that he underwent his first psychoanalysis, with Sándor Ferenczi. The dual nature of Róheim's training made it possible for him to do pioneering work as the first psychoanalytic anthropologist. His work in the early 1920s clearly reveals his unusual conceptual equipment. In 1921 he delivered a paper on Australian totemism (published in an extended version in 1925), for which he received the Freud prize in applied psychoanalysis. In the same year, he published a paper on "Das Selbst" ("The Self") that anticipated psychoanalytic ego psychology by almost two decades. Again, Róheim's treatise "Nach dem Tode des Urvaters" (1923) reveals his application of anthropological research to Freudian concepts. Although he accepted Freud's postulation of the primal horde and the Cyclopean family, Róheim insisted that the "eating of the overlord" represented nothing more than the reversion to a fantasy pertaining to the oral stage of human development. This theory foreshadowed his shift from Freud's phylogenetic explanation of culture (in terms of the sense of guilt and the experience of sin and atonement, arising from Oedipal conflicts during the phallic stage) to an ontogenetic one.

With the encouragement of Freud himself, and assisted by Princess Marie Bonaparte, another of Freud's disciples, Róheim sought to find anthropological evidence to validate Freud's theories and set out in 1928 on the first field expedition undertaken by a psychoanalytically trained anthropologist. His research was particularly focused on the western tribes of central Australia (Aranda, Pitche-tara, Pindupi, Yumu, Nambutji). His stay in central Australia was followed by a nine-month visit to Sipupu, Normanby Island, in the D'Entrecasteaux group. His aim was to study a matrilineal society closely resembling that of the Trobrianders, whom Malinowski had made the focus of his field

observations. Róheim also visited briefly among the Yuma Indians on the borders of California, Arizona, and New Mexico. It was his research among the Australian aborigines, however, that provided the basis for his revised psychoanalytic theories.

Freud's conception of the relationship between human culture (particularly totemism, religion, and social structure) and the vicissitudes of the primal family has found little confirmation in biological or anthropological evidence. Some evidence for the Cyclopean family, as conceived by Freud, may be adduced from the life of anthropoids, especially that of the baboon horde (Zuckerman 1932), but it must be remembered that this study was made of apes in captivity. Although the primal horde theory does seem to be useful in the interpretation of numerous myths and religious practices, serious doubt as to the validity of the hypothesis that memories may be inherited led Róheim to base his field work and his clinical observations on the theory that cultural patterns and institutions can be explained in large measure by the biologically conditioned crucial role of the specific infantile situation, such as separation anxiety, together with the largely environmentally conditioned prevailing libidinal trend.

Róheim thus proposed an ontogenetic theory of culture, with special emphasis on the role of biology and anatomy, that depended heavily on the findings of the Dutch anatomist Louis Bolk (1926). According to Bolk, comparative morphology reveals that the human individual shows many traits of neoteny, such as hairlessness, the form of the ear muscles, Mongoloid development, orthognathia, the central position of the *foramen magnum*, relative brain weight, certain variations of the jaw, persistent cranial sutures, and the tendency toward brachycephalization. However, among the higher anthropoids, the eruption of the milk teeth starts almost immediately after birth, and shortly after the growth of the second milk molar the first permanent one appears. This rather rapid change demands an equally rapid development of the jaw as well as of the entire skull. In man, on the other hand, there are two intervals that impose retardation. The milk teeth are fully grown only toward the end of the second year, so that the human child depends on sucking rather than on biting for a much longer period. After the second year there is an interval of about four years before the first permanent molars come through.

The psychological consequences of neoteny, according to Róheim and additional explorations by Muensterberger, stem chiefly from man's prolonged

dependence on maternal care. The so-called ego functions reflect the long period of human infancy; indeed, the various psychological processes do not reach maturity until the second decade of life. The human child, more than any other mammal, is in constant need of protection and nourishment provided by an external agency. The bond created by physiological necessity develops into a necessary emotional and social tie, forcing the human being to attain and maintain continuous relationships with other people. Like several other psychoanalysts, Róheim considered this dual bond to be a peculiarly human predicament. Moreover, this protracted symbiotic relationship forms the nucleus of wider social cohesion and organization. In its negative aspect of dependence and jealousy, it is the source of feelings of anxiety and helplessness and of fears of castration and separation.

This uniquely human dual bond leads to the differentiation of ego and nonego, of a self and an external world. It produces man's predominant aim, which is to gain contentment and happiness and to avoid discontent and disillusionment. Attachment to the mother is the first source of pleasure, and disillusionment and anxiety are created by her absence. Thus, from a psychobiological point of view, the human organism can discharge tension only through another organism; the infant is anaclitic. The dual bond also leads to the differentiation of ego and id, which are essentially one in lower animals. "Instincts" (built-in behavior patterns) are replaced by learned behavior patterns.

Róheim's dream theory is constructed on the analogy of the infantile experience of initial acceptance of the outside world as only an intermediary aim on the path of withdrawal from that world. His emphasis on the significance of particular fantasies experienced during the process of falling asleep is related to his ontogenetic approach. He related the sense of *falling* that is characteristic of the hypnagogic state to the inherent desire to return to the dual-unity situation that denies the separation of the child from the mother. He suggested that in sleep we return to the intrauterine situation and that dreams are attempts to re-establish contact with reality. In other words, dreams are efforts to reverse the regressive condition caused by sleep and therefore constitute a defense against the reinfantilization represented by sleep.

The ontogenetic theory of culture explains man's pattern of separation (anxiety, ambivalence, aggression) and subsequent reunion as a repetition of the fundamental duality of the mother-child relationship. The reaction to absence or frustra-

tion is aggression and/or anxiety. The pattern also appears in numerous primitive ceremonies and religious rituals, including the killing and oral introjection of the totem animal, behavior which, according to Róheim, represents aggression as a reaction to separation anxiety, followed by reunion or the covenant (1945). Similarly, the Oedipus complex is not the result of inherited memory images; it is the inevitable outcome of the human family and its extended period of infancy. Totemism, in this view, is merely one of the most frequently used of the several available solutions to the Oedipal nature of man.

Using his own field observations as examples, Róheim was able to show the development of specific cultural patterns or group ideals. Among the aborigines of central Australia, the typical infantile trauma is the *alknarintja* situation: the mother lies on top of the boy infant, ostensibly to protect the child. Evidence from myths, dreams, and rituals suggests that the child experiences this "protection" as a threat to himself when he is in a helpless, passive, inverted position. The institution of phallic ceremonialism, based on the exclusion of women and the projection of the threatening mother image onto the dangerous phallic female demons, is an attempt to ward off the early traumatization of the boy infant. Thus, this simple example from Pitche-tara society shows how the crucial elements in the infantile situation lead to reaction formations, which then become institutionalized. The smaller and more self-sufficient a society, the greater the likelihood that specific cultural patterns in the relationship between mother and child will produce a prevalent personality type in the society, that is, a greater psychological similarity among the members of that society.

As Róheim put it: "Man has invented culture because of his delayed infancy or intolerance of tension. In our own attempt to master reality we create a society, i.e., a well-functioning symbiotic mode of existence. Instead of parricide and incest, we have the super-ego and group formation, instead of 'clutchings' to archaic objects we have 'seekings' of new substitute objects. Adaptation then becomes a favorable solution between narcissism and object-cathexis" (1943, pp. 77-80). Hence, in Róheim's view, culture is a defense system against the fear of object loss and the separation anxiety that sacrifices immediate gratification and the monopolization of maternal love for a modicum of security and partial gratification.

WARNER MUENSTERBERGER AND BILL DOMHOFF



[For the historical context and subsequent development of Róheim's ideas, see CULTURE AND PERSONALITY and MYTH AND SYMBOL.]

## WORKS BY RÓHEIM

- 1911 Sárkányok és sárkányölő hősök (Dragons and Dragon Killers). *Ethnographia* (Budapest) 22:129-142, 193-209.
- 1921 Das Selbst. *Imago* 7:1-39.
- 1923 Nach dem Tode des Urvaters. *Imago* 9:83-121.
- 1925 *Australian Totemism: A Psycho-analytic Study in Anthropology*. London: Allen & Unwin.
- 1941 Die psychoanalytische Deutung des Kulturbegriffs. *Internationale Zeitschrift für Psychoanalyse und Imago* 26:9-31.
- 1943 *The Origin and Function of Culture*. Monograph No. 6. New York: Nervous and Mental Disease Monographs.
- 1945 *War, Crime and the Covenant*. Monticello, N.Y.: Medical Journal Press.
- 1952 *The Gates of the Dream*. New York: International Universities Press.

## SUPPLEMENTARY BIBLIOGRAPHY

- BOLK, LOUIS 1926 *Das Problem der Menschwerdung*. Jena (Germany): Fischer.
- MUENSTERBERGER, WARNER 1955 On the Biopsychological Determinants of Social Life. *Psychoanalysis and the Social Sciences* 4:7-25.
- ZUCKERMAN, SOLLY 1932 *The Social Life of Monkeys and Apes*. New York: Harcourt.

## ROLE

## I. PSYCHOLOGICAL ASPECTS

Theodore R. Sarbin

## II. SOCIOLOGICAL ASPECTS

Ralph H. Turner

## I

## PSYCHOLOGICAL ASPECTS

The concept of role came into psychology through the cross-disciplinary efforts of a number of social and clinical psychologists. The dramatic metaphor was first systematically exploited by George H. Mead and the University of Chicago sociologists in the 1920s. Its utility was demonstrated particularly in connection with the analysis of occupational and quasi-occupational categories. In the 1930s and 1940s several writers opened the way for social scientists to employ the concept in the analysis of interpersonal behavior. Among them were Moreno (1934), who pioneered the use of role playing as a tool of psychotherapy and role training; Sarbin (1943), who analyzed hypnotic behavior as role enactment; Cameron (1947), who demonstrated how the behavior of disordered persons could be understood as faulty role enactment, as failure in taking the role-of-the-other, and as ineptness in playing out one's socially assigned roles; and Newcomb (1950),

who used role as a central concept in his social psychology. Several writers have characterized role as the construct that bridges psychology and the other social sciences. It connotes not only overt actions and performances but also covert expectations held by an observer, or by a group of observers, such expectations serving as the basis for judging the propriety of the enactment. The concept of role is now widely used in the study of industrial and bureaucratic organizations, the family, small-group interaction, behavior modification, attitude change, behavior pathology, and recently, the analysis of psychological experiments where persons are assigned the role of experimental subject.

**Definition of role.** In general, the term "role" continues to be used to represent the behavior expected of the occupant of a given position or status. Thus, following the implication of the dramaturgical metaphor, an actor assigned to the position (or part) of Hamlet is expected to enact the role of Hamlet, the role being characterized by certain actions and qualities. A person who is assigned to the position of clergyman (or who elects to be placed in such a position in the social structure) is similarly expected to enact the role of clergyman characterized by certain typical actions and qualities. In this definition two features are emphasized: (1) *expectations* (i.e., beliefs, cognitions) held by certain persons in regard to what behaviors are appropriate for the occupant of a given position, and (2) *enactments* (i.e., conduct) of a person who is assigned to, or elects to enter, a given position.

## Role enactment

Although the definition given above is satisfactory as a first approximation to the analysis of ongoing social behavior, more refined analysis has demanded the introduction of concepts that are more differentiated. In the first place, the behavior that serves as the dependent, or outcome, variable is seen as *role enactment*. That is, interest is focused on what the occupant of a given position *does and says*, such behaviors being noted by observers who may use behavior rating scales or free response verbalizations to make known their observations. Needless to say, the effectiveness, convincingness, validity, or propriety of such role enactments varies between persons and situations. To uncover the antecedents of this variation, social scientists have introduced additional theoretical conceptions and observational procedures. Among the variables that have been demonstrated to be antecedent to variation in effectiveness or

propriety of role enactment are (1) validity of role expectations held by the actor, (2) accuracy of the actor in locating the other(s) (and reciprocally the self) in the proper role system, (3) sensitivity to situationally generated role demands, (4) available general and specific skills, (5) congruence of self and role, and (6) reinforcement properties of the audience. Each of the variables will be discussed below.

**Role expectations.** In order effectively to perform, i.e., to validate his occupancy of a social position, the actor must learn—either at first hand or vicariously—what performances are associated with what positions. In short, he must be acquainted with at least certain sectors of the social system: he must know what obligations, privileges, rights, and duties are the defining characteristics of each position he may be called on to occupy. Given such knowledge, he is in a better position to enact an assigned or ascribed role. The concept of role expectation is implicit in many practices in contemporary society. College students, for example, are frequently participants in orientation programs designed to provide them with the proper role expectations. Induction into the military or into bureaucratic organizations is similarly designed to acquaint inductees with their rights and privileges and with their duties and obligations. Failure to acquire expectations for recurring roles leads to enactments that are judged as inept, invalid, improper, antisocial, or illegal.

Role expectations may be viewed as *actions* or *qualities* expected of the occupant of a position. If viewed as actions, role expectations are codified as in a job description: the occupant is expected, for instance, to call the roll, open the windows, secure the doors. If viewed as qualities, role expectations are codified in adjectival terms; for example, the occupant is expected to be warm, friendly, outgoing, sincere, and cautious. This account of role expectations is necessarily oversimplified. Persons usually occupy multiple positions, some of which may be considered compatible, such as professor, father, and scientist, and some considered incompatible, such as high school teacher and bartender. The resolution of such incompatibilities has been studied in a number of contexts. These studies have provided us with a recognition of institutionalized rituals and personal adaptive techniques that serve to reduce the strain attendant upon occupying incompatible roles. In addition, the occupant of a position not only enacts role behavior that is congruent with the expectations held by others in the macrocosmic social system but also performs acts that make good his occu-

pancy of concurrent positions in microcosmic personal role systems. Whereas the macrosystem points to the kinds of role behaviors appropriate to regularly occurring situations, miniature systems develop where the role expectations must be expressed in terms separate from the more obvious task expectations of the macrosystem. These may be denoted as idiosyncratic preferences, nonconsensual expectations, and subtle or disguised expectations. Goffman (1961) has provided some enlightening illustrations of this notion under the rubric "encounters." In a surgical team, for example, the chief surgeon may not only direct the flow of activities that are task oriented, thus meeting the expectations for surgeon, but also engage in behavior unrelated to the surgical task—such as making jokes—designed to maintain an optimal degree of tension in the performers.

The problems of assessing role expectations are no different from the problems of assessing any inferred variable. Expectations must be inferred from overt conduct, including self-reports, and from samples of behavior in unstructured test situations. An investigator may assess the role expectations held by a group of persons by observing their conduct in situations where an actor emits behavior in response to standard problems. Or he may make his inferences from replies to more direct questions, such as "What behavior should one typically expect from a person who is in the position of school superintendent?" Various devices have been created for more systematic investigation of role expectations, such as role inventories, designed to get at the *actions* expected of the occupant of a position, and adjective check lists and rating scales, designed to get at the *qualities* or personal dispositions expected of the occupant (Sarbin & Jones 1955).

**Role perception.** It is axiomatic that in order to survive as a member of a society, a person must be able to locate himself accurately in the role structure. The simplest way to accomplish this is by seeking and finding answers to the question "Who am I?" Since roles are structured in reciprocal fashion, the answers can also be achieved through locating the position of the other by implicitly asking the question "Who are you?" The answers to the latter question are usually phrased in terms of role categories, such as man, teacher, friend, officer, secretary, and clown. In order to establish the position of the other, the actor must pay attention to the behaviors emitted by that other, scanning for cues that have reliability and validity. Physique and figure, hairstyle, facial adornment, and dress, among other things, are



cues to which the actor may attend in order to locate the other, and reciprocally the self, in social space. The process of enculturation is heavily weighted with the learning of which emitted behaviors are signs or cues for which positions in the social structure. In ambiguous or partially structured situations, the actor may influence the role of the other by casting himself in one rather than another permitted role. His behavior, in short, serves as a potential constraint on the role enactment of the other, who must also locate himself in the social structure.

Under ordinary conditions the location of the other in social space occurs without much reflection, partly because much social behavior is ritualized, partly because the "badges of office" are more or less obvious. To locate the position of the other it is sometimes necessary to infer covert behavior on the part of the other, such as "feelings," expectations, and attitudes. The more obvious cues and badges of office may not provide necessary information for correctly classifying the other. In this case the actor may search the ecology for additional cues, such as might be emitted in the form of posture, linguistic forms, facial expression, and gesture. Illustrative empirical research arising from this conception are two studies carried out in an effort to establish the relationship between accuracy of role perception on the basis of postural cues and indices of deviant or maladjusted role enactment. Sarbin and Hardyck (1955) constructed 43 stick figures in which all cues except postural ones were eliminated. These stick figures, black-and-white line drawings schematically depicting human postures, are presented one at a time to the subject, who selects, from among 5 choices, a word or phrase descriptive of the stimulus object. On standardizing groups, the frequencies of response to the 5 choices follow a J curve, the modal response containing at least 70 per cent of the distribution. When a subject's protocol for the 43 stick figures shows a high proportion of modal responses, his performance is designated as *conformant*. Inmates of a state mental hospital who were classified as schizophrenic (whose deviant role enactments presumably were instrumental in establishing residence in the hospital) selected the modal response choice no more often than the low response choices.

A further study (Krasner et al. 1964) made use of a more refined scoring system. Those stick figures whose modal responses were terms indicating *covert* states, such as thinking, sadness, and anger, were scored apart from those stimuli whose modal responses were terms representing

more overt conduct, such as running, sleeping, and standing at attention. Comparison of the performances of nonhospitalized control groups with hospital patients classified as schizophrenic showed significant differences in scores for the inference items but no significant differences for the non-inference items. This finding lends support to the proposition that the conduct of persons classified as exhibiting deviant role enactment (i.e., mental patients) is characterized by errors in locating others in the role system when they must depend on interpreting cues that represent covert states or dispositions.

Stated in an alternate form, the role perception variable considers the cue properties of the ecology, especially cues arising from the behavior of other persons. The accuracy with which a person notices such cues and draws conclusions regarding the role of the other is directly related to the accuracy with which he locates himself in the reciprocal or complementary role. The validity, propriety, and convincingness of his role enactment, then, depends on how accurately he perceives the role of the other. [See PERCEPTION, articles on PERSON PERCEPTION and SOCIAL PERCEPTION.]

**Role demands.** Once a person correctly locates the position of the other on the basis of cues emitted by the other interactant, the possible choice of role behaviors is reduced from near-infinity to a small number. Further constraints on the choice of role behaviors are introduced when certain additional features of the situation are taken into account. These may be called role demands, that is, demands for a specific role enactment.

Silently operating to guide the actor in his choice of roles, these demands stem from cultural norms and may, in fact, dictate a role choice that is contradictory to requirements under conditions when the demands are not operative. Among such demands one may list norms for modesty, communication, control of aggression, cooperation, face-saving, public commitment, and prevention of embarrassment to others. Violation of such norms ordinarily carries heavy sanctions against the violator, hence their power to constrain the choice of role behavior.

Two series of experiments by Orne (1959) illustrate in a striking way the necessity of taking into account the role demand variable, or, to use his designation, the demand character of the environment. In his first series of studies he demonstrated that the role enactments of subjects properly motivated to simulate hypnotic subjects could not be differentiated from the role enactments of subjects who were instructed through traditional

hypnotic procedures. By experimentally manipulating the role demands of the simulators, he approximated the demand character of the situation of the genuine hypnotic subjects. His simulators were instructed by an associate to pretend to be hypnotized and to try to deceive an expert in hypnotism. This study, together with a series of studies reported by Sarbin (1950), raises the possibility that hypnotic behavior is role enactment and that the manipulated demand for a specific role enactment is in large measure responsible for the outcome. One of the demands used by professional entertainers who perform hypnosis is that embarrassment to the other, in this case the hypnotist, be prevented. In a public setting the subject is in a position to embarrass the hypnotist by rejecting his instructions. However, for most persons the prevention of embarrassment is a powerful norm, and the informed hypnotist will capitalize on this fact.

Orne's second series of studies (Orne & Scheibe 1964) is addressed to the social aspects of the psychological experiment. Generalizing from the effects of manipulating role demands in the hypnosis experiment, he hypothesized that the outcomes in psychological research might also be related to implicit or explicit demands. Such demands operate when a subject wittingly or unwittingly learns that he is in the experimental group or in the control group. The experiment was conducted in a laboratory setting that simulated sensory deprivation experiments, with an examination by a white-coated physician and the presence of a panic button in the experimental room, but without actual sensory deprivation. The experimental subjects reported many, if not all, of the bizarre and extreme phenomena attributed to actual sensory deprivation. Subjects who were in the control condition received the same experimental treatment (without the panic button) and reported none of the sensory deprivation phenomena. In brief, different role demands led to different outcomes. In this experiment the assignment to experimental or control group was made explicit. In the usual psychological experiment the subject is left free to speculate on whether he is in the experimental or control group. He will use whatever inputs are available to help define his role. The outcome of his performance in the role of experimental subject will be influenced by his responsiveness to the intended or unintended role demands. [See HYPNOSIS; PERSUASION; SUGGESTION.]

**Role-taking aptitude.** Cameron (1947), Gough (1948), and Sarbin (1950), among others, have presented arguments for the existence of a skill,

aptitude, or competence that facilitates role enactment. Presumably acquired early in life through subtle learning experiences, this role-taking aptitude may be regarded as parallel to other aptitudes, such as mechanical, clerical, and drawing aptitudes. Authorities have not always been clear in their use of the term "role-taking aptitude"; in some contexts the referent is the ease with which a person enacts a number of roles, as in the case of the successful professional actor; in others the referent is the facility with which a person adopts and enacts a role on the basis of limited information. In the former referent the point of interest is role enactment as motoric (including gestural and verbal) behavior; in the latter it is cognitive and perceptual behavior.

Empirical studies have focused primarily on the cognitive component of this skill. With few exceptions, there has been no accurate measurement of the skill as a general cognitive competence. Most investigators, in fact, have drawn the conclusion that accuracy in predicting the behavior of others is a function of "stereotype accuracy"; i.e., the prediction is accurate insofar as the target person is regarded as an exemplar of a class of persons of known characteristics. If the subject knows the role system and the expectations that make up this system, he can extrapolate to any individual occupying a position in the system. The work of Gage (1953) is typical. Predictions of conduct were more accurate when based on general role expectations than on perceived idiosyncrasies of the target person. Although the experimental literature warns us to remain skeptical about a general role-taking aptitude, Cline and Richards (1960) appear to have demonstrated that such a disposition does exist. In a series of carefully controlled experiments some subjects made verifiable inferences about the social behavior of target persons after observing samples of behavior in sound motion pictures. The accuracy of these inferences could not be attributed to such conceptions as stereotype accuracy.

The motoric component has not been systematically studied except in the context of dramatic acting. Like the cognitive component, the motoric component of role-taking skill presupposes broad acquaintance with, and practice of, overt role behaviors associated with many positions. Although every person's repertory of verbal and gestural actions is different from every other's, a minimum of this skill is required for social interaction. Assessment of this variable can at present be achieved only through judgments made by observers. For example, Sarbin and Lim (1963) had their experi-



mental subjects perform pantomime improvisations before a panel of specialists in dramatic arts who recorded on a rating scale their impressions of the convincingness of the performance. The purpose of the experiment was to determine whether relative responsiveness in the hypnosis situation could be attributed to the motoric component of the role-taking skill. The findings clearly showed that part of the variation in hypnotic behavior can be attributed to the motoric role-taking skill. In earlier writings the role-taking skill was likened to empathy, or the ability to take the part of the other. The term implied that the actor would take the perspective of the target person, sharing not only his attitudes and beliefs but also his feelings.

Critics of role theory have pointed to the confidence man as a person who is gifted in taking the role of the other but who obviously does not genuinely share the perspective of the victim. This observation has been used to question the utility of the concept of role-taking aptitude. The criticism may be outlined somewhat as follows: If the confidence man really takes the role of the other, would he not then see himself as the victim, feel the pain of loss or embarrassment, and reverse his behavior? This criticism confuses two dimensions in the cognitive process. The first is the basis of knowledge of people in general, acquaintance with the class of persons of which the intended victim is a member, and study of the intended victim himself, the confidence man draws inferences about the probable role behavior of the victim, given certain role demands that the confidence man may manipulate. The second dimension is an evaluative one and has no necessary connection with the first. The value that one places on his own conduct need bear no more relation to his skill in role taking than to any other skill.

**Self-role congruence.** The writings of George H. Mead contain observations and speculations about two conceptions that have become thoroughly entrenched in modern psychology, role and self. The extensions of role have already been discussed. For a more complete analysis of the variation in effectiveness of role enactment, the concept of self must be invoked. However, it does not stand alone but must be regarded as one of a set of conceptions—this set describable on a dimension of congruence or compatibility. The belief that a role is performed better if its requirements are consonant with one's "natural inclinations" comes to us from folk psychology. By translating "natural inclinations" to self concept, or self characteristics, the

folk belief can be translated into a scientifically testable hypothesis, namely, that the effectiveness or validity of role enactment is related to the degree of congruence of self and role.

In brief, the self may be defined as the residues of a human being's transactions with objects and events, including other people. These residues are the referents for the symbol *I* (Sarbin 1954). To assess self structures, a large number of procedures have been devised, including autobiographies, speech and handwriting samples, responses to unstructured inkblots, endorsements of personality test items, and adjective check lists. In our linguistic system the most frequently used device for talking about self characteristics is the adjective. A person can describe his self as a catalogue of traits through the use of lists of adjectives, such as friendly, quiet, calm, vain, confident. The adjective check list, widely used in research settings, is based on this observation. Another widely used procedure in contemporary psychology is the personality inventory. This instrument is made up of incomplete *I*-statements requiring the respondent to indicate various predicates with himself as the subject, in this way expressing his preferences, attitudes, feelings, beliefs, values, and/or opinions. The dimensions used for organizing such sentences into scales are constructed for particular purposes by the investigator or distilled through analytic and statistical procedures. [See PERSONALITY MEASUREMENT, article on PERSONALITY INVENTORIES; SELF CONCEPT.]

Through the use of adjective check lists and inventories of *I*-statements one can systematically specify the qualitative requirements of a role (for example, a club treasurer must be honest and efficient) and compare such requirements with the self characteristics of persons who are assigned the role (for example, George sees himself as honest and efficient). When self and role characteristics are congruent, a convincing performance is expected, other things being equal. When self and role characteristics are incongruent or antagonistic, the enactment is likely to be invalid and unconvincing.

A study reported by Smelser (1961) illustrates the use of the variable of self-role congruence. Two samples of experimental subjects were selected, the members of one group having high scores on the trait "dominance," derived from the California Psychological Inventory, the members of the second having low scores (see Gough 1957). The subjects, in pairs, were placed in a cooperative work situation and were systematically assigned

roles calling for dominant or submissive enactments. The most productive pairs were those where dominant subjects had been assigned the dominant role and nondominant subjects the submissive role. The least productive pairs were those where the self characteristics and role characteristics had been reversed.

Employment of the self-role congruence variable has been especially useful in predicting effective enactment in the hypnosis role. Sarbin and Andersen (1964) have reported experiments whose results could be interpreted as showing that persons whose self characteristics are similar to the requirements of the role of hypnosis subject enact that role with greater convincingness. Their procedure involved the construction of an inventory of I-statements whose items reflected the kinds of role behaviors traditionally associated with hypnosis. Two of the dimensions of the role are (1) to become absorbed, to concentrate, and (2) to accept readily altered attentional and mood experiences. Illustrative of the I-statements written to tap corresponding self characteristics are the following statements: "I sometimes become so absorbed in a task that I am forgetful of other, less important aspects of the daily routine"; "I would not be afraid of anything that would temporarily change my awareness of my surroundings." The study yielded significant correlations between scores on the self-assessment inventory and the degree of convincingness in the enactment of the role.

An alternate way of regarding this variable is available. Where the self becomes involved in the role, the enactment is more likely to be judged as valid, convincing, and proper. It has been suggested that forced compliance to a role whose requirements are incompatible with one's self characteristics will not produce changes in the self. For example, attempts to indoctrinate prisoners of war in alien ideology generally fail because of lack of self-involvement in the role, even though the prisoners may give the appearance of involvement through the use of histrionic talents. [See BRAIN-WASHING.]

**Reinforcement properties of the audience.** Unlike many current attempts to understand social psychological phenomena, role theory addresses itself to *continuities* in conduct rather than to small samples of outcome behaviors. Social roles are enacted over time: the performances are protracted. To employ the dramaturgical metaphor further, roles are enacted before audiences—sometimes a one-person audience, as in psychoanalytic therapy;

sometimes a small group audience, as in a work or discussion group; sometimes a large mass audience, as at political meetings, the theater, and professional football games; and sometimes a symbolic audience, as in a writer's imagined picture of his reading audience. Audiences may serve two functions, providing cues helpful to the actor in locating his role and providing social reinforcements or sanctions. The plethora of studies on operant conditioning may be taken as a paradigm for the reinforcement properties of the audience (Krasner & Ullman 1965). In operant conditioning, selected responses of the subject are reinforced through social reinforcements, such as approval, praise, eye contact, friendly gesture, and nodding, or through more direct reinforcements, such as food and money. In the usual case of social interaction, audiences primarily provide social reinforcements. The effect of positive reinforcements, other things being equal, is to encourage the actor to continue his performance and to think that it is convincing, valid, and proper. Negative reinforcements, on the other hand, communicate to the actor that his role enactment is invalid and discourages the continuation of the role. [See COMMUNICATION, MASS, especially the article on AUDIENCES; LEARNING, articles on INSTRUMENTAL LEARNING and REINFORCEMENT.]

This article has attempted to identify current conceptions that may be organized under the general heading of role theory. The outcome variable of interest to psychologists whose work is guided by this orientation is *role enactment*, that is, the overt conduct of a person in his effort to validate the occupancy of the positions to which he is assigned in various macrostructures and microstructures. Role enactments are judged by audiences along various dimensions, in general reflecting the convincingness, propriety, effectiveness, or validity of the enactment. Because role enactments vary in these dimensions, investigators have attempted to isolate and identify the sources of variation. Six components of effective role enactment have been described and at least one empirical study introduced to illustrate the utility of each variable. The validity, effectiveness, or convincingness of role enactment depends on (1) a valid set of role expectations, (2) an accurate assessment of the role of the other (and reciprocally the self), (3) sensitivity to the role demands of the situation, (4) role-relevant cognitive and motoric aptitudes, (5) self characteristics congruent with the role, and (6) positive reinforcements for appropriate



acts. The six variables are listed as if they were independent of one another. As yet they can be designated only as conceptually independent; further research will tell to what extent they are empirically independent—and under what conditions.

THEODORE R. SARBIN

[Other relevant material may be found in PERCEPTION, articles on PERSON PERCEPTION and SOCIAL PERCEPTION; SOCIOMETRY; STATUS, SOCIAL; STRATIFICATION, SOCIAL; SYMPATHY AND EMPATHY; and in the biography of MEAD.]

#### BIBLIOGRAPHY

- CAMERON, NORMAN A. 1947 *The Psychology of Behavior Disorders: A Biosocial Interpretation*. Boston: Houghton Mifflin.
- CLINE, VICTOR B.; and RICHARDS, JAMES M. JR. 1960 Accuracy of Interpersonal Perception: A General Trait? *Journal of Abnormal and Social Psychology* 60:1-7.
- GAGE, NATHANIEL L. 1953 Accuracy of Social Perception and Effectiveness in Interpersonal Relationships. *Journal of Personality* 22:128-141.
- GOFFMAN, ERVING 1961 *Encounters: Two Studies in the Sociology of Interaction*. Indianapolis, Ind.: Bobbs-Merrill.
- GOUGH, HARRISON G. 1948 A Sociological Theory of Psychopathy. *American Journal of Sociology* 53:359-366.
- GOUGH, HARRISON G. 1957 *The California Psychological Inventory: Manual*. Palo Alto, Calif.: Consulting Psychologists Press.
- KRASNER, LEONARD; and ULLMAN, LEONARD P. 1965 *Research in Behavior Modification: New Developments and Implications*. New York: Holt.
- KRASNER, L.; ULLMAN, L.; and WEISS, R. L. 1964 Studies in Role Perception. *Journal of General Psychology* 71:367-371.
- MORENO, JACOB L. (1934) 1953 *Who Shall Survive? Foundations of Sociometry, Group Psychotherapy and Sociodrama*. Rev. & enl. ed. Beacon, N.Y.: Beacon House.
- NEWCOMB, THEODORE M. 1950 *Social Psychology*. New York: Dryden.
- ORNE, MARTIN T. 1959 The Nature of Hypnosis: Artifact and Essence. *Journal of Abnormal and Social Psychology* 58:277-299.
- ORNE, MARTIN T.; and SCHEIBE, KARL E. 1964 The Contribution of Nondeprivation Factors in the Production of Sensory Deprivation Effects: The Psychology of the "Panic Button." *Journal of Abnormal and Social Psychology* 68:3-12.
- SARBIN, THEODORE R. 1943 The Concept of Role-taking. *Sociometry* 6:273-285.
- SARBIN, THEODORE R. 1950 Contributions to Role-taking Theory: I. Hypnotic Behavior. *Psychological Review* 57:255-270.
- SARBIN, THEODORE R. 1954 Role Theory. Volume 1, pages 223-258 in Gardner Lindzey (editor), *Handbook of Social Psychology*. Cambridge, Mass.: Addison-Wesley.
- SARBIN, THEODORE R.; and ANDERSEN, M. L. 1964 Base Rate Expectancies and Motoric Alterations in Hypnosis. *International Journal of Clinical and Experimental Hypnosis* 12:147-158.

SARBIN, THEODORE R.; and HARDYCK, CURTIS D. 1955 Conformance in Role Perception as a Personality Variable. *Journal of Consulting Psychology* 19:109-111.

SARBIN, THEODORE R.; and JONES, D. S. 1955 An Experimental Analysis of Role Behavior. *Journal of Abnormal and Social Psychology* 51:236-241.

SARBIN, THEODORE R.; and LIM, DONALD T. 1963 Some Evidence in Support of the Role-taking Hypothesis in Hypnosis. *International Journal of Clinical and Experimental Hypnosis* 11:98-103.

SMELSER, WILLIAM T. 1961 Dominance as a Factor in Achievement and Perception in Cooperative Problem Solving Interactions. *Journal of Abnormal and Social Psychology* 62:535-542.

#### II

#### SOCIOLOGICAL ASPECTS

The concept of *role*, borrowed from the stage, has been central in those sociological analyses which seek to link the functioning of the social order with the characteristics and behavior of the individuals who make it up. The concept has a relatively long history in American sociology and has attracted interest from some of its best minds. For instance, Park (1926, p. 137) noted that "everyone is always and everywhere, more or less consciously, playing a role. . . . It is in these roles that we know each other; it is in these roles that we know ourselves." In this and most subsequent usages, the following elements appear in the definition of role: it provides a comprehensive *pattern* for behavior and attitudes; it constitutes a *strategy* for coping with a recurrent type of situation; it is *socially identified*, more or less clearly, as an entity; it is subject to being played recognizably by *different individuals*; and it supplies a major basis for *identifying* and *placing* persons in society.

George H. Mead's observation (1934) that in mature social behavior the individual works out his own role by imaginatively taking the role of the other soon led to the idea that articulation between the roles played by partners in interaction determines whether interaction is harmonious and productive. The theme that personal and social disorganization arises from failure of the individual to choose between incompatible roles was developed by Mowrer (1935), and the problem of balancing rights and obligations in role articulation was explored by Kirkpatrick (1936). In the same year, Linton (1936) presented the first systematic statement identifying role as a *segment of culture*; he held that it consisted of behavioral prescriptions or norms bearing one-to-one correspondence with social status [see STATUS, SOCIAL; the biography of LINTON]. All these traditions have inspired descrip-

tive studies, stressing personal and organizational strains from incomplete articulation of roles.

More recently such topics as the spontaneous emergence of similar informal roles in different discussion groups (Benne & Sheats 1948) and the hypothesis of role specialization in task and social-emotional behavior (Bales 1950) have been the subject of experimental study. Many organizational studies have demonstrated that lack of clarity and consensus in role conceptions is a factor in reducing organizational effectiveness and morale (see, for instance, Gross et al. 1958). One serious effort has been made to review and appraise this body of literature comprehensively (Rocheblave-Spenlé 1962).

This variety of studies is an index of some major differences in usage. A few writers equate "role" with the actual *behavior* of an individual, although most distinguish between role as expected behavior or conceptions and role behavior as enactment. Furthermore, there is a continuing issue of generality, raised by Linton's restriction of role to the prescriptions associated with a status (1936). Another influence is that of behaviorism, which led to an effort to think of a role as an inventory of specific behaviors or conditioned responses (Cottrell 1942), rather than as certain purposes, sentiments, and strategies in the context of which both routinized and improvised behavior are mixed. Authorities also differ on whether "role" should be said to consist only of the *norms* applicable to a position or should include not only norms but also expectations and conceptions with respect to behaviors that are only in the faintest degree normative. Finally, role is defined by some as essentially a part to be learned and played, whereas for others it is primarily a way of coping with an imputed other-role.

### Emergence of roles

Despite definitional differences, it is possible to state, as a series of rough "postulates," the tendencies that are usually assumed to constitute the normal role processes. Role theory will first be considered as it would apply to a self-contained, dynamic system of interacting roles; considerations of organizational and societal setting, as well as of individual personality, will be discussed later.

**Role differentiation and accretion.** In any interactive situation, behavior, sentiments, and motives tend to be differentiated into units, which can be called roles; once roles are differentiated, elements of behavior, sentiment, and motives that appear in the same situation tend to be assigned to the existing roles. This postulate suggests that one should

study degrees of differentiation, as indicated by *clarity* of role definition and degree of *role overlap*. The hypothesis has been widely accepted that individual adjustment to role and group effectiveness is positively related to clarity and negatively related to overlap, although excessive clarity may be overly restrictive to the individual and rigidifying to the organization, and some overlap may facilitate empathic understanding and contribute to interpersonal bonds. The postulate also suggests that attention should be given to what is grouped and what is separated into different roles. Internal consistency of roles is thought to facilitate individual adjustment and group efficiency; internal contradiction is held to be explained with the help of such concepts as role set (discussed below), culture contact and assimilation, imposition of role structure by authority, and rigidification through formalization and tradition. Feasibility of enactment by a single individual has been the most common principle invoked to explain the peculiar way in which the content of one role can be separated from that of another; additional principles, such as those arising from the functional requirements of the group, are noted below.

**Meaningfulness.** In any interactive situation the meaning of individual actions for ego (the actor) and for any alters (partners in interaction) is assigned on the basis of the imputed role. This postulate is inspired by the common observation that identical actions mean totally different things when viewed as expressions of different roles.

**Behavioral correspondence.** A tendency (never completely realized) for the social character of a role to shift toward correspondence with patterns of behavior is actually seen to take place within the role context. For example, changes in the wife and mother role have been attributed to removal of many traditional domestic activities from the home, despite inertial factors that caused a lag in the adaptation.

**Evaluation.** Every role tends to acquire an evaluation, both in terms of rank and in terms of its socially favorable or detrimental character. The perceived utility of the role, the power resident in the role, and association of the role with other valued roles are among hypothesized determinants of role evaluation.

### Role as interactive framework

**Interaction.** The establishment and persistence of interaction tend to depend on the emergence and identification of ego and alter roles. This postulate extends the meaningfulness postulate to



interaction. Identifications are tentative and are based on cues, so the character of interaction in any initial encounter, or under grossly changed conditions, is tentative and changeable.

**Role complementarity.** Each role tends to form as a comprehensive way of coping with one or more relevant alter roles. The actual character of any given role involves a constant strain toward equilibrium between actual behavior, evaluative adjustments between roles, and the tension of maintaining each as a viable role in relation to the other. For example, in any role relationship of superordination-subordination, such as parent-child, teacher-student, master-servant, policeman-criminal, the subordinate role will incorporate mechanisms that afford some latitude for action.

**Role consensus.** There is a tendency toward consensus regarding the content of roles in interaction. Because of the adaptation that takes place as roles form, some consensus regarding what enactors of each role are to do seems inevitable, although the degree of consensus required for a viable interaction system is not great. Role dissensus has generally been thought an obstacle to harmonious interaction (as in the family), although the evidence is not entirely consistent.

**Legitimate expectation.** There is a tendency for stabilized roles to be assigned the character of legitimate expectations, implying that deviation from expectation is a breach of rules or violation of trust. Willard Waller's observations (1938) about the sense of injustice that arises in married life when one partner deviates from what has come to be the customary pattern of behavior suggest the following hypothesis: the degree to which ego ascribes legitimacy to his expectations for alter's role varies directly with the degree to which the selection and enactment of his role depend on alter's role and role behavior. The evaluation of the role also appears to make a difference, suggesting this principle: to the extent to which alter's role is positively valued with respect to rank and favorableness, ego will translate his anticipations regarding alter's role into legitimate expectations. Thus an unfavorably evaluated role, such as criminal or indigent, evokes legitimate expectations of continuance in the role only to the extent of justifying adaptive behavior by relevant others, such as surveillance by police and anticipatory defensive behavior.

### **Role and actor**

**Role persistence.** Once stabilized, the role structure tends to persist, regardless of changes in the

actors. This postulate takes two specific forms. First, when an actor leaves the group and is replaced by another, there is a tendency to allocate to the new member the role played by the one who leaves. Second, if one actor changes roles, there is a tendency for another actor to make a compensatory change of roles in order to maintain the original role structure. The latter principle has been referred to as role appropriation by Perry, Silber, and Bloch (1956), who noted that in some families, when the parent became disorganized and assumed a childlike role of dependence in a disaster situation, a child suddenly blossomed into responsibility and helped to supply family leadership.

**Role allocation.** There is a tendency for a given individual to be identified with a given role, and a complementary tendency for an individual to adopt a given role, for the duration of the interaction. The postulate of role allocation, in combination with the meaningfulness postulate, suggests that an actor's behavior is without clear meaning until he settles on a particular role in interaction. In combination with the postulate of legitimate expectation, role allocation suggests that an individual whose behavior cannot be fitted into an identifiable role will be regarded as violating the terms of interaction.

A role allocation has taken place when relevant alters interact with ego on the basis of the same role as he is playing. Ego's act of attempting to determine the role that alter will play has been dubbed "altercasting" by Weinstein and Deutschberger (1963). The alter's use of a relatively unambiguous response to confer the right to play the chosen role is sometimes designated as *validation*.

The bases for role allocation can be inferred from the postulate of evaluation in conjunction with the postulate of role complementarity (see above). Since roles are evaluated, there is likely to be competition for the preferred roles, and the principles of any competitive situation will be at work. Because of complementarity of roles, it is essential that an individual be able to play the role with sufficient adequacy; demonstrated adequacy or inadequacy therefore affects the willingness of others to allocate to him a particular role.

**Role taking.** To the extent to which ego's role is an adaptation to alter's role, it incorporates some conception of alter's role. If the actor is guided chiefly by his conception of alter's role (role taking), his behavior has an element of improvisation about it; he simply follows the course of action that seems reasonable for coping with alter under

the circumstances. A further implication of the role-taking postulate is that the individual learns roles in pairs or sets, not singly.

**Role adequacy.** Role behavior tends to be judged as "adequate" or "inadequate" by comparison with some definite conception of the role in question. When combined with the role allocation postulate, this observation leads to the following generalization: to the degree to which a role is favorably evaluated, judgments of role adequacy tend to be translated into evaluations of the person playing the role.

**Role reciprocity.** The degree of role adequacy legitimately expected of ego tends to be a function of the degree of role adequacy attributed to alter. Conversely, the degree to which ego can legitimately claim the privileges of his role tends to be a function of his degree of role adequacy. With this postulate, which has been elaborated by Gouldner (1960), we call attention to two facts, namely, that role behavior does not normally correspond precisely with role conceptions, and that legitimate expectations are adjusted in the course of interaction.

### Organizational setting

The organizational setting supplies both direction and constraint to the working of the processes already outlined, thus bringing them into more complex interrelationships.

**Organizational goal dominance.** To the extent to which roles are incorporated into an organizational setting, organizational goals tend to become the crucial criteria for role differentiation, evaluation, complementarity, legitimacy of expectation, consensus, allocation, and judgments of adequacy. To the extent to which the organization has well-defined goals, the most salient role differentiation will be according to different kinds of tasks that together accomplish the goal. The expectations assigned greatest legitimacy and held with most consensus will tend, under these circumstances, to be those with most obvious task relevance; judgments of role adequacy will therefore weigh task adequacy more heavily than other aspects of role adequacy.

**Legitimate role definers.** To the extent to which roles are incorporated into an organizational setting, the rights to define the legitimate character of roles, to set the evaluations on roles, to allocate roles, and to judge role adequacy tend to be lodged in particular roles.

**Status.** To the extent to which roles are incorporated into an organizational setting, role differ-

entiation tends to link roles to statuses in the organization.

**Role set.** To the extent to which roles are incorporated into an organizational setting, each tends to develop as a pattern of adaptation to multiple alter roles. The teacher role must incorporate tenable adaptations to pupils, parents, other teachers, and principals, as well as to less salient alters. Discrepant expectations may arise from relations with each of the alters; Merton (1957) describes several mechanisms that are employed to minimize the conflict in the role set.

**Formalization.** To the extent to which roles are incorporated into an organizational setting, their persistence is intensified through tradition and formalization. A strain then develops between formalized or traditional role definitions and the informal role structure.

### Societal setting

Because society is built on accommodation among many organizations, it introduces multiple organizational referents for roles and multiple roles for the actor.

**Economy of roles.** Similar roles in different contexts tend to become merged, so as to be identified as a single role recurring in different relationships. As a corollary, roles in situations of limited generality and social significance tend to be shaped in accordance with the pattern of roles in situations of greater generality and social significance. One index of this tendency is that we often speak of male and female roles, of heroic and unheroic roles, of judge and defendant, when seeking meaning and order in quite simple interaction.

**Value anchorage.** To the extent to which roles are referred to a societal context, differentiation tends to link roles to social values. Status is not thereby made less important as a point of anchorage, but social values take the place of organizational goals as major anchorages for roles. Values thus supply one of the major bases for the simplification of an otherwise unmanageable profusion of recognized roles in society; even in more limited contexts, as we have seen in the discussion of role economy, roles for interaction tend to be fitted into a framework of social values.

**Allocation consistency.** The individual in society tends to be assigned and to assume roles that are consistent with one another. The role conflict which arises from failure of this process has been subject to considerable study. Mechanisms such as organizational compartmentalization facilitate the tendency; there are also various recognized transi-



tional cues, as when a functionary who wishes to pass from one role to another speaks of "putting on a different hat." Perhaps the most specific mechanism for achieving consistency is the tendency for roles that have the greatest generality, such as age, sex, and occupation, to serve as qualifying criteria for the allocation of other roles.

### Role and person

When we turn to the consequences of role process for the individual as a sociological entity, we look at the *actor* in relation to a role-in-situation, and we observe the *person* in the setting of society. For the actor the primary concern is coping with strain; for the person the dynamics hinge on the management of the several roles in his repertoire.

**Role strain.** The actor tends to act so as to alleviate role strain arising from role contradiction, role conflict, and role inadequacy and to heighten the gratifications of high role adequacy. Role strain comes largely from failure of the many processes already outlined to function adequately, so as to leave unclear, incomplete, and contradictory elements in a role. This may come about through failure of role cues, gross lack of role consensus, and so forth. Little has been noted that is distinctive about adjustments to role strain, as compared to ways of coping with any other type of stress.

**Socialization.** The individual in society tends to adopt a repertoire of role relationships as a framework for his own behavior and as a perspective for the interpretation of the behavior of others. It is important to remember that socialization is partly a matter of learning somewhat culturally standardized role conceptions and partly a matter of acquiring a viable image of the social world by means of the tentative and adaptive processes that make up role enactment.

**Self-conception.** The individual tends to form a self-conception by selective identification of certain roles as more characteristically himself than other roles in his repertoire. The self-conception, from a sociological point of view, is not a set of traits but an organization of roles that both articulates the person with the society and incorporates role evaluations. On the basis of the roles included, the individual is said to develop a sense of personal prestige, which is likely to be reflected in his bearing, his self-assurance, and other aspects of his interpersonal relations. On the basis of his perceptions of role adequacy in these most "ego-involved" roles (Sherif & Cantril 1947), the individual develops a self-esteem.

**Adaptivity of self-conception.** The self-conception tends to stress those roles which supply the

basis for effective adaptation to relevant alters. Thus, while there is no simple or uniform self-conception held by all persons defined by society as "criminal," the established criminal must somehow take account, in his self-conception, of the fact that his criminal identification is recurrently the most salient consideration shaping the behavior of others toward him in a variety of situations.

**Role distance.** To the extent to which the person must play roles that contradict his self-conception, these roles will be assigned "role distance," together with mechanisms for demonstrating the lack of personal involvement that the person feels when playing these roles (Goffman 1961).

RALPH H. TURNER

[Directly related are the entries IDENTITY, PSYCHOSOCIAL; INTERACTION; NORMS; STATUS, SOCIAL. Other relevant material may be found in PERSONALITY MEASUREMENT, article on SITUATIONAL TESTS; SOCIAL PSYCHOLOGY; SOCIALIZATION; and in the biographies of LINTON; MEAD; PARK; WALLER.]

### BIBLIOGRAPHY

- BALES, ROBERT F. 1950 *Interaction Process Analysis: A Method for the Study of Small Groups*. Reading, Mass.: Addison-Wesley.
- BANTON, MICHAEL P. 1965 *Roles: An Introduction to the Study of Social Relations*. London: Tavistock.
- BENNE, KENNETH D.; and SHEATS, PAUL 1948 Functional Roles of Group Members. *Journal of Social Issues* 4, no. 2:41-49.
- BIDDLE, BRUCE J.; and THOMAS, EDWIN J. (editors) 1966 *Role Theory: Concepts and Research*. New York: Wiley.
- COTTRELL, LEONARD S. JR. 1942 The Adjustment of the Individual to His Age and Sex Roles. *American Sociological Review* 7:617-620.
- GOFFMAN, ERVING 1961 *Encounters: Two Studies in the Sociology of Interaction*. Indianapolis, Ind.: Bobbs-Merrill.
- GOODE, WILLIAM J. 1960 A Theory of Role Strain. *American Sociological Review* 25:483-496.
- GOULDNER, ALVIN W. 1960 The Norm of Reciprocity: A Preliminary Statement. *American Sociological Review* 25:161-178.
- GROSS, NEAL; MASON, WARD S.; and MCEACHERN, ALEXANDER W. 1958 *Explorations in Role Analysis: Studies of the School Superintendency Role*. New York: Wiley.
- KIRKPATRICK, CLIFFORD 1936 The Measurement of Ethical Inconsistency in Marriage. *Ethics* 46:444-460.
- LINTON, RALPH 1936 *The Study of Man: An Introduction*. New York: Appleton.
- MEAD, GEORGE H. 1934 *Mind, Self and Society From the Standpoint of a Social Behaviorist*. Edited by Charles W. Morris. Univ. of Chicago Press.
- MERTON, ROBERT K. 1957 The Role-set: Problems in Sociological Theory. *British Journal of Sociology* 8:106-120.
- MOMENT, DAVID; and ZALEZNIK, ABRAHAM 1963 *Role Development and Interpersonal Competence: An Experimental Study of Role Performance in Problem-*

- solving Groups. Boston: Harvard Univ., Graduate School of Business Administration, Division of Research.
- MOWRER, HARRIET 1935 *Personality Adjustment and Domestic Discord*. New York: American Book.
- NEIMAN, LIONEL J.; and HUGHES, JAMES W. 1951 *The Problem of the Concept of Role: A Re-survey of the Literature*. *Social Forces* 30:141-149.
- PARK, ROBERT E. 1926 *Behind Our Masks*. Survey (New York) 56:135-139.
- PARSONS, TALCOTT 1951 *The Social System*. Glencoe, Ill.: Free Press.
- PERRY, STEWART E.; SILBER, EARLE; and BLOCH, DONALD A. 1956 *The Child and His Family in Disaster: A Study of the 1953 Vicksburg Tornado*. National Research Council, Committee on Disaster Studies, Disaster Study No. 5. Washington: National Academy of Sciences-National Research Council.
- ROCHEBLAVE-SPENLÉ, ANNE MARIE 1962 *La notion de rôle en psychologie sociale: Étude historico-critique*. Paris: Presses Universitaires de France.
- ROMMETVEIT, RAGNAR (1953) 1955 *Social Norms and Roles: Explorations in the Psychology of Enduring Social Pressures With Empirical Contributions From Inquiries Into Religious Attitudes and Sex Roles of Adolescents From Some Districts in Western Norway*. Oslo: Akademisk Forlag; Minneapolis: Univ. of Minnesota Press.
- SHERIF, MUZAHER; and CANTRIL, HADLEY 1947 *The Psychology of Ego-involvements, Social Attitudes and Identifications*. New York: Wiley; London: Chapman & Hall.
- SLATER, PHILIP E. 1955 *Role Differentiation in Small Groups*. *American Sociological Review* 20:300-310.
- STRODTBECK, FRED L.; and MANN, RICHARD D. 1956 *Sex Role Differentiation in Jury Deliberations*. *Sociometry* 19:3-11.
- WALLER, WILLARD (1938) 1951 *The Family: A Dynamic Interpretation*. Revised by Reuben Hill. New York: Dryden.
- WEINSTEIN, EUGENE A.; and DEUTSCHBERGER, PAUL 1963 *Some Dimensions of Altercasting*. *Sociometry* 26:454-466.
- ZNANIECKI, FLORIAN 1965 *Social Relations and Social Roles: The Unfinished Systematic Sociology*. San Francisco: Chandler. → Published posthumously.

## ROMAN CATHOLICISM

See CHRISTIANITY; and the articles under RELIGION.

## RORSCHACH, HERMANN

Hermann Rorschach, Swiss psychologist, was born in 1884 and died in 1922. These were the exciting years when psychoanalysis came of age and gave impetus to fresh inquiry in all social and cultural areas. In the psychology of the individual, psychoanalysis revealed hitherto unfathomed aspects of the emotions. Scientific psychology was thus faced with a new challenge, the measurement of unconscious emotions. Rorschach's famous test met this newly felt need.

Rorschach, the first of three children, was born in Zurich. When he was two years old the family moved to Schaffhausen, a commercially busy and culturally stimulating city; his father became a drawing teacher there. His mother died when he was 12 and his father when he was 18. In his final year in the *Kantonschule*, undecided whether to choose a career in science or art, he wrote to the biologist Ernst Haeckel (O. Rorschach 1944; Morgenthauer 1921), who advised him to go into medicine. He began his studies in Zurich and went successively to Berlin, to Bonn, then back to the University of Zurich, where he took his degree in 1912. His dissertation on hallucinations forecast his interest in psychopathology. The research was done under Eugen Bleuler's supervision.

In 1910 Rorschach married Olga Stemplin, a Russian, who surely had something to do with his decision, in 1913, to take a post in a sanitarium near Moscow. It was his third journey to Russia, and Olga Rorschach has described (1944) his favorable reaction to the vastness of that country, to its people, its churches, and its landscape. He had learned Russian, and he wrote a novel (unpublished) about Pushkin and later planned a work on Dostoevski. After less than a year in Russia, he returned to Switzerland. He had held posts in three mental hospitals when he was appointed as assistant director in the cantonal institute in Herisau, and he remained in this position until he died of acute peritonitis in 1922. His one important professional office was that of vice-president, in 1919, of the then newly organized Swiss Psychoanalytic Society; Emil Oberholzer was the president. Some correspondence with Morgenthauer (1965) shows how strenuously Rorschach promoted psychoanalysis in Swiss medical circles.

Rorschach was undoubtedly supported in his psychoanalytic orientation by Bleuler, who was among the first to receive Freud's ideas sympathetically. Jung, a resident in a suburb of Zurich, was giving impetus to experimentation with free association as well as to extraversion-introversion theory. Bleuler's psychiatry and the theories of Jung and Freud were the three principal sources of the ideas that were to crystallize in Rorschach's test. He used psychoanalytic theory in interpreting the test record of one of Oberholzer's patients (Rorschach & Oberholzer 1923) and also in an article on two Swiss leaders of religious sects (1927), whom he had studied before he developed the test. In this paper he traced personality changes from the introverted to the extraverted type. He anticipated his own exposition in the *Psychodiagnostics* ([1921] 1942, pp. 102 ff.).



Rorschach's test clearly reflects his particular scientific preferences, powerful as the influences of Freud, Jung, and Bleuler may have been. His basic method was empirical and his temperament that of the experimentalist. Rorschach himself so described his work in the brief foreword to the *Psychodiagnostics*: "The entire work is predominantly empirical in character" (1921). He offered suggestions for controlled experiments, especially with regard to those two critical variables—color, indicative of acted out feelings, and movement, indicative of ideas lived in the imagination, such as wish fulfillments or daydreams ([1921] 1942, p. 58). They are the main ingredients of his *Erlebnistypus*, the core of his personality concept. The ratio between the number of movement responses and the weighted value of color responses is an index of the *Erlebnistypus*. The ratio can change, as can also the *Erlebnistypus*, with the increase or reduction of either color or movement responses.

Rorschach's broad psychological interests are reflected in his penetrating studies of some Swiss religious sects. According to Morgenthaler (1965), he was striving for a synthesis of folkways, ethnology, religion, and psychopathology in a single sociological system. Some of his most important work was done on schizophrenia.

It is intriguing to apply his personality theories to what is known of Rorschach himself. Warmth of feeling for others was one of his traits, one that he acquired chiefly from his mother, a kindly and sociable person (Ellenberger 1954). To his father he owed his aptitude for drawing, and he encouraged his patients to draw. He had many acquaintances but only a few intimate friends. According to his own *Erlebnistypus* concept, he would thus be introversion. However, Olga Rorschach writes, "He was constantly growing, ever expanding his *Erlebnistypus*, from introversion towards extraversion. He thus attained to an enviable ambiequality" (1944). And she quotes Rorschach as saying that "the ideal is the ambiequal."

Rorschach was a venturesome explorer, and in that most taxing terrain, the human personality. While he clung to the observable, he arrived at his conclusions more by his insights than by the logic of statistics. His short life did not permit him to complete his work. But as a tool for clinical exploration as well as for general research in personality, Rorschach's test has gone far toward meeting the challenge faced by a science of psychology.

SAMUEL J. BECK

[For the historical context of Rorschach's work, see PSYCHOANALYSIS and the biographies of BLEULER;

FREUD; JUNG. For discussion of the subsequent development of Rorschach's ideas, see FANTASY; PROJECTIVE METHODS, especially the article on THE ROSCHACH TEST.]

#### WORKS BY ROSCHACH

- (1921) 1942 *Psychodiagnostics*. 3d ed. Bern: Heber; New York: Grune. → First published in German.
- (1923) 1924 ROSCHACH, HERMANN; and OBERHOLZER, E. The Application of the Interpretation of Form to Psychoanalysis. *Journal of Nervous and Mental Disease* 60:225-248; 359-379. → First published in German as "Zur Auswertung des Formdeutversuches für die Psychoanalyse." Reprinted in the 1942 edition of Rorschach 1921.
- 1927 Zwei schweizerische Sektenstifter. *Imago* 13:395-441.
- 1965 *Gesammelte Aufsätze*. Bern: Huber.

#### SUPPLEMENTARY BIBLIOGRAPHY

- BECK SAMUEL J. 1963 Rorschach's *Erlebnistypus*: An Empiric Datum. *Schweizerische Zeitschrift für Psychologie und ihre Anwendungen*, Beiheft 45:8-25.
- ELLENBERGER, HENRI 1954 The Life and Work of Hermann Rorschach (1884-1922). *Menninger Clinic, Bulletin* 18:173-219.
- LOOSLI-USTERI, MARGUERITE 1956 Adresse Présidentiale. *Schweizerische Zeitschrift für Psychologie und ihre Anwendungen*, Beiheft 34:15-22.
- MORGENTHAUER, W. (1921) 1942 Biographical Sketch. Page 9 in Hermann Rorschach, *Psychodiagnostics*. 3d ed. Bern: Heber; New York: Grune.
- MORGENTHAUER, W. 1965 Erinnerungen an Hermann Rorschach. Pages 95-101 in Hermann Rorschach, *Gesammelte Aufsätze*. Edited by K. W. Bash. Bern: Huber.
- ROSCHACH, OLGA 1944 Über das Leben und die Wesensart von Hermann Rorschach. *Schweizerisches Archiv für Neurologie und Psychiatrie* 53:1-11.

#### ROSCHER, WILHELM

Wilhelm Georg Friedrich Roscher (1817-1894), German economist, was a descendant of an old Hanoverian family of officials and judges. He studied history and political science at Göttingen and Berlin. In 1840 he became a lecturer in both these subjects at the University of Göttingen, where in 1843 he rose to the rank of professor-extraordinary (comparable to associate professor), and the following year to full professor. In 1848 he was called to the University of Leipzig, where he remained until his death.

Roscher is commonly considered, with Karl Knies and Bruno Hildebrand, to be one of the founders of the "older historical school" of German economics. In his small book *Grundriss zu Vorlesungen über die Staatswirtschaft nach geschichtlicher Methode* (1843) he based the study of political economy on the principles of historical investigation. He was one of the first economists

other than Friedrich List to have done so; although in France Auguste Comte had argued in his *Cours de philosophie positive*, 1830–1842, that the historical method should be applied to all the social sciences, he had himself concentrated on sociology.

The *Grundriss* was the basis for Roscher's widely read *System der Volkswirtschaft*, a work of five volumes published between 1854 and 1894 and described in its subtitle as a handbook and reader for businessmen and students. It became the most influential textbook of political economy in Germany in the second half of the nineteenth century. Roscher also wrote some purely historical and philosophical books and articles in his early years, some political treatises, and many studies in the history of economic literature. His only work of lasting significance, however, is his *Geschichte der National-oekonomik in Deutschland* (1874).

Roscher called his method "historical" or "historico-physiological," in contrast to the "philosophical" or "idealistic" method. He contended that the object of political economy is not to establish the best possible state of things, but to describe the actual state at which the economy has arrived through continual development. Of the two crucial questions, "What is?" (positive economics) and "What ought to be?" (normative economics), Roscher wanted to answer only the former; the latter question, in his view, could not be studied scientifically because it depends on constantly changing opinions and individual choices. While a science could never be based on the German idealistic philosophy, the study of what actually occurred could provide a "firm island of scientific truth which may be accepted in the same manner as the adherents of different systems of medicine all admit the teaching of mathematical physics." ([1854–1894] 1901–1922, vol. 1, p. 80).

Compared to the "younger historical school," led by Gustav Schmoller, Roscher was much closer to the classical theory of economics. In contrast to List, whose focus was on the productive forces of every nation and who was therefore interested only in the dynamics of development, Roscher accepted the value theory of the classical school for the analysis of any given stage of the economy and conceived of a process of organic development taking place between statically conceived stages. He thought that the life of nations, like the vegetable and animal world, has four such stages of development: childhood, youth, manhood, and old age. Thus, development is in both an ascending and a descending direction (in contrast to List's theory that evolution is exclusively ascending).

Roscher believed that development could best be

studied by inquiring into the history of ancient nations like Greece and Rome, because that history has been terminated. Three factors of production govern the evolution of every nation: land, labor, and capital. Of these, one is always predominant in each successive stage of life. A close, and especially a statistical, knowledge of economic facts, according to the "objectivist" spirit of Leopold von Ranke, would permit, in Roscher's view, the solution of conflicts of interest between the factors of production. Roscher was much more skilled in collecting obscure details and understanding particular events and theories than in formulating a theory of his own. Lacking any systematic theory, he was not forced to come to terms with the historical facts he collected or to recognize the futility of his search for "laws of nature" or "laws of evolution."

Roscher could not fulfill the declared aim of his historical approach: to achieve for political economy what the historical approach of F. K. von Savigny and Karl F. Eichhorn had already achieved for the field of law and legislation. However, his *Geschichte der National-oekonomik in Deutschland* remains an outstanding work that must be consulted by everyone who does research in that field. Roscher was interested not only in the history of economic analysis but also in the personality of every economist and in the character of every economic work. His book is, therefore, a compendium, perhaps even an encyclopedia, of the history of German economists and of German economic and economic-political literature up to 1874.

EDGAR SALIN

[For the historical context of Roscher's work, see ECONOMIC THOUGHT, article on the HISTORICAL SCHOOL, and the biographies of LIST; RANKE; SAVIGNY. For discussion of the subsequent development of Roscher's ideas, see the biographies of HILDEBRAND; KNIES; SCHMOLLER.]

#### WORKS BY ROSCHER

- 1843 *Grundriss zu Vorlesungen über die Staatswirtschaft nach geschichtlicher Methode*. Göttingen (Germany): Dietrich.
- (1854–1894) 1901–1922 *System der Volkswirtschaft: Ein Hand- und Lesebuch für Geschäftsmänner und Studierende*. 5 vols. Stuttgart (Germany): Cotta. → Volume 1: *Die Grundlagen der Nationalökonomie*, 1922. Volume 2: *Nationalökonomik des Ackerbaues und der verwandten Urproduktionen*, 1912. Volume 3: *Nationalökonomik des Handels und Gewerbetreibenden*, 1917. Volume 4: *System der Finanzwissenschaft*, 1901. Volume 5: *System der Armenpflege und Armenpolitik*, 1906.
- (1861) 1878 *Ansichten der Volkswirtschaft aus dem geschichtlichen Standpunkte*. Heidelberg: Winter.
- (1874) 1924 *Geschichte der National-oekonomik in Deutschland*. 2d ed. Munich and Berlin: Oldenbourg.



(1895) 1917 *Geistliche Gedanken eines Nationalökonomens*. New ed. Dresden (Germany): Zahn & Jaensch.

#### SUPPLEMENTARY BIBLIOGRAPHY

- CUNNINGHAM, WILLIAM 1894 Why Had Roscher So Little Influence in England? *American Academy of Political and Social Science, Annals*. 5:317-334.
- ONCKEN, A. (1910) 1963 William Roscher: 1817-1894. Volume 3, pages 323-327 in Robert H. I. Palgrave, *Dictionary of Political Economy*. Rev. ed. New York: Kelley Reprints.
- SCHMOLLER, GUSTAV 1888 Wilhelm Roscher. Pages 147-171 in *Zur Literaturgeschichte der Staats- und Sozialwissenschaften*. Leipzig: Duncker & Humblot.
- Verzeichnis von W. Roschers Schriften in der Originalsprache und in Uebersetzungen. 1894 Sächsische Akademie der Wissenschaften zu Leipzig, Philologische-Historische Klasse, *Berichte*. 46:222-226. → A comprehensive bibliography of Roscher's writings.
- WEBER, MAX (1922) 1951 Roschers historische Methode. Pages 3-42 in *Gesammelte Aufsätze zur Wissenschaftslehre*. 2d ed. Tübingen (Germany): Mohr.

### ROSS, EDWARD A.

Edward Alsworth Ross (1866-1951) was one of the founders of American sociology and is perhaps best remembered for his militant advocacy of melioristic sociology—a sociology dedicated to the cause of social reform. He was internationally known, also, as a sociological theorist.

Born of Scotch-Irish parents in Illinois, Ross grew up on the Middle Border: as a child, he was taken to Centralia, Kansas, and then to Cedar Rapids, Iowa. Orphaned when he was not yet ten years old, he was cared for by various relatives in Iowa and finally given a foster home with the Beach family in Marion, Iowa; he maintained close ties with Mrs. Beach until her death. At the age of 16 Ross left Marion to study at the preparatory school of Coe College, and 4½ years later he received his B.A. from that institution. Following his graduation, he studied in Germany and at Johns Hopkins, where he received his Ph.D. in 1891. After teaching for one year at Indiana University and then at Cornell, Ross joined the faculty at Stanford University in 1893 as a full professor of administration and finance.

Ross's fervent Populist-Progressive views on controversial social issues soon aroused the wrath of California conservatives. Mrs. Leland Stanford, the widow of the founder of Stanford University, repeatedly urged David Starr Jordan, the president of that university, to dismiss Ross; Jordan interceded for Ross on several occasions but was finally forced to demand his resignation. In November 1900 Ross accompanied his enforced resignation with a public statement blaming Mrs. Stanford for

his ouster. Seven of his colleagues subsequently resigned in sympathy or were similarly dismissed. Throughout the country, Ross's dismissal became a *cause célèbre*. Public figures and publicists censured the university for dismissing a scholar because of his political views, and Ross, in the role of martyr, encouraged the public outcry. He continued his academic career first at the University of Nebraska and then, in 1906, at the University of Wisconsin, where he taught until his retirement in 1937.

The sources of the various components of Ross's thought, of his sociology, his Progressivism, and his intermittent nativism, are to be found in the social and economic tensions of a predominantly agrarian society in the process of urbanization. Reared as he had been in the moralistic Middle West, Ross's response to this disruptive transformation was to seek to retain the values of nineteenth-century America: the dignity of the individual, personal and public morality, and the continuing progress of man and society. The virtues of the past would survive if public opinion were guided by enlightened leaders, if legal sanctions were applied in a sophisticated manner, and if supernatural religion were abandoned. Thus Ross sought to synthesize the old and the new, to infuse an impersonal industrial society with the idealized virtues of the face-to-face community in which he grew up.

Ross's important contributions to American sociology came early in his career. His sociological theory was allied to his ideology: in his *Foundations of Sociology* (1897-1904), a collection of essays, he attacked conservative social Darwinism and tried instead to fashion a sociological theory that could be used to understand and reform American society. He applied his erudition to the demolition of such outmoded theories as the monism of Buckle, Loria, and Vico; the bio-organicism of von Lillienfeld and Schäffle; the genetic evolutionism of DeGreef and L. H. Morgan; and, with particular determination, he assaulted the mechanistic theory of Herbert Spencer. Ross did accept two basic nineteenth-century concepts: organicism and positivism. However, he altered organicism by stripping it of its biological and physical implications and redefining it in social and psychological terms. He refocused the positivist approach on the study of social phenomena in small units, as the necessary basis for a valid science of society. If Ross did not succeed in making organicism completely social or positivism completely operational, he must nevertheless be credited with communicating the need for a new definition of the subject matter and methodology of sociology.

*Foundations* was an important work not only because it presented this new definition but also because it was the first book in American sociology to stress the importance of social processes as a sociological concept. Ross divided the social processes—the general and recurring phenomena that pervade the entire social order—into major categories and subcategories. In 1905, he described 11 major categories (for example, cooperation and competition) and 32 subcategories; by 1920, he had reduced the major divisions to four (association, domination, exploitation, and opposition). However, Ross never arrived at a definitive number of processes, nor did he ever establish the exact nomenclature to describe them. He saw these processes as accounting for such phenomena as the power of the army, the structure of the family, the functioning of the church, and the nature of the government. The analysis of social processes constituted the core of the textbooks he wrote between 1920 and 1940, and he utilized them to explain the civilizations of the past as well as contemporary society: war, for example, is the product of “competition,” slavery the result of “segregation” and “subordination,” and democracy the consequence of “equalization.”

Ross's sociological formalism—evidenced by his study of social processes—was supplemented by research on social behavior. *Social Control* (1901) and *Social Psychology* (1908) delineated the formal and informal ways in which society constrains the behavior of the individual. *Social Control* contained Ross's finest contribution to sociological thought and brought him international fame. Like William Graham Sumner's *Folkways* (1906) and Charles Horton Cooley's *Human Nature and the Social Order* (1902), *Social Control* was an attempt to explain order. Sumner understood social stability as the evolving “cultural system” produced by folkways and mores; Cooley stressed the role played by the subjective processes in the socialization of the individual (a “psychological system”) in creating stability. Taking a position between the two, Ross presented a “social system” that is both individualistic and interactional in that it integrates to some degree the external norms emphasized by Sumner and the inward processes discussed by Cooley. In addition to discussing the ways and means of societal control of the individual, Ross dealt with what he called the “grounds” of control, that is, the conditions for the need of social control, and with the relationship of social change to the techniques of control.

*Social Control* is still valuable as a listing of the means society has to control the individual: custom

and convention, legal and social sanction, religion and education. But the operation of social control is more complicated than it seemed to Ross in 1901. In his later years Ross recognized this and made an unsuccessful effort to remedy this failing; he never published this piece of work. He was unable to explain in a manner satisfactory to modern, predominantly interactionist, social psychologists the mechanism by which external norms (custom and convention, for example) are internalized.

Ross's *Social Psychology* (1908) is likewise obsolete. It was based on the Tardean imitation-suggestion theory and remained an academic best seller even when the Tardean theory itself was no longer accepted. Ross did not incorporate interactionist theories in his book, and this eventually led to its identification with an outmoded phase of social psychology that was concerned with mobs and panics, custom and convention. Nevertheless, *Social Psychology* served to stimulate an interest in that discipline and pointed to the need in the academic curriculum for the separate study of social psychology.

As a sociologist Ross faithfully served the cause of Progressivism, that outgrowth of the middle-class reform spirit which pervaded American life and thought in the first 15 years of the twentieth century. In *Foundations*, his definition of the social processes mirrors the temper and values of Progressive America. In *Social Control* and in *Social Psychology*, Ross was preoccupied with the problem that confronted all Progressives: the preservation of both individual freedom and social stability in a mass society. Ross did not see social stability threatened by the breakdown of social tradition and of supernatural religion, as did European social psychologists like Le Bon and Sighele. As a Progressive, he welcomed the dissolution of traditional institutions and rituals and even condoned an occasional mob action in the cause of social justice.

Throughout his career Ross was interested in practical reforms both at home and abroad, and his *Social Control* and *Social Psychology* are filled with specific suggestions for the mitigation of social evils. He did not hesitate to urge political leaders in foreign countries on every continent (he visited them all) to adopt American ways: to industrialize their economies and to democratize their political systems.

In addition to constructive Progressive ideas, a minor note of nativism was evident in Ross's thought for many decades. Until 1930 he was a spokesman for Anglo-Saxon superiority. His nativist syndrome included opposition to continued Japanese immigration to the United States in the



1890s and to southern and eastern European immigration in the early decades of the twentieth century, a preoccupation with eugenics, and an obsession with differential fertility rates. His nativism reveals his limited faith in the power of social institutions to cure social evils. Reform needed the reinforcement of nativism if the virtues of nineteenth-century rural American life were to be restored or perpetuated.

Ross was physically imposing (he was 6½ feet tall) and personally attractive. He could be both rigidly moral and shrewdly supple. He was generous but not effusive, frugal but not penurious. Despite his Stanford reputation and two subsequent incidents involving academic freedom at Wisconsin, he was far from seeking entanglement in controversies. He took a rather detached view of institutions and organizations, being reluctant either to exercise authority or to be dependent upon the authority of others. To be sure, he served two terms as president of the American Sociological Society (1914 and 1915) and was chairman of the sociology department at the University of Wisconsin from 1929 until 1937. With the exception of Lester F. Ward, whom Ross regarded intellectually and emotionally as a father figure, he rarely became involved in profound friendships or in taxing personal feuds, thus leaving himself free to pursue a productive as well as a lucrative career. (His introductory textbooks *The Principles of Sociology*, published in 1920, and *New-age Sociology*, published in 1940, were widely used.)

Ross is still a figure of consequence to his profession. As a reformer and, for a time, a creative social theorist, he deserves respect; as a writer and popularizer of sociology he has had few peers in the annals of American sociology.

JULIUS WEINBERG

[For the historical context of Ross's work, see SOCIAL DARWINISM and the biographies of COOLEY; LORIE; MORGAN, LEWIS HENRY; SPENCER; SUMNER; VICO; WARD, LESTER F. For discussion of the subsequent development of Ross's ideas, see SOCIAL CONTROL.]

#### WORKS BY ROSS

- (1897-1904) 1920 *Foundations of Sociology*. 5th ed. New York: Macmillan. → A collection of Ross's essays and lectures at the turn of the century; first published in book form in 1905. Especially valuable as a survey of contemporary sociological theory at an important juncture in the history of American sociology.
- (1901) 1928 *Social Control: A Survey of the Foundations of Order*. New York and London: Macmillan. → Dealing with the problem of how American mass society can achieve a balance between social stability and

individual freedom, this book represents Ross's most creative contribution to sociological thought.

- 1907 *Sin and Society: An Analysis of Latter-day Iniquity*. Boston: Houghton Mifflin. → A moralistic Progressive tract that makes a sophisticated plea for sociological jurisprudence—a demand that society bring its legal theory and practice abreast of current problems.
- (1908) 1918 *Social Psychology: An Outline and Source-book*. New York: Macmillan. → An exposition of Ross's Tardean imitation-suggestion theory and a primer in social psychology on which a generation of students was raised.
- (1909) 1919 *Changing America: Studies in Contemporary Society*. New York: Century. → A collection of Progressive pieces with popular appeal.
- 1919 *What Is America?* New York: Century. → Originally written for the U.S. government during World War I as a piece of "interpretation" (propaganda) to be used for overseas consumption and later converted by Ross for domestic readers. Extensively used in translation by foreign-language newspapers as part of the Americanization program after the war. An excellent summary of Ross's Progressive outlook on American history and society.
- (1920) 1938 *The Principles of Sociology*. 3d ed. New York: Century. → A mature presentation of Ross's sociological theory and findings. A highly readable text used for many years throughout the country.
- 1922 *The Social Trend*. New York: Century. → A collection of previously published essays and lectures on current popular issues.
- (1924-1927) 1928 *World Drift*. New York: Century. → An anthology of articles and addresses on foreign and domestic affairs.
- 1936 *Seventy Years of It: An Autobiography*. New York: Appleton. → Boastful and vain, yet an informative, revealing autobiography.
- 1940 *New-age Sociology*. New York and London: Appleton. → A further revision of the earlier texts and a not-too-successful attempt to keep data and theory abreast of current sociological thought.

#### SUPPLEMENTARY BIBLIOGRAPHY

- COOLEY, CHARLES H. (1902) 1956 *Human Nature and the Social Order*. Rev. ed. In Charles H. Cooley, *Two Major Works: Social Organization and Human Nature and the Social Order*. Glencoe, Ill.: Free Press. → Each title reprinted with individual title page and pagination. Separate paperback editions were published in 1964 by Schocken.
- HERTZLER, JOYCE O. 1951 Edward Alsworth Ross: Sociological Pioneer and Interpreter. *American Sociological Review* 16:597-613.
- HOUSE, FLOYD NELSON 1936 *The Development of Sociology*. New York: McGraw-Hill.
- KARPF, FAY BERGER 1932 *American Social Psychology: Its Origins, Development, and European Background*. New York: McGraw-Hill.
- KOLB, WILLIAM L. 1948 The Sociological Theories of Edward Alsworth Ross. Pages 819-832 in Harry Elmer Barnes (editor), *An Introduction to the History of Sociology*. Univ. of Chicago Press.
- MARTINDALE, DON 1960 *The Nature and Types of Sociological Theory*. Boston: Houghton Mifflin.
- SUMNER, WILLIAM G. (1906) 1959 *Folkways: A Study of the Sociological Importance of Usages, Manners, Customs, Mores, and Morals*. New York: Dover.

## ROUSSEAU, JEAN JACQUES

In most of his writings, not merely in the *Contrat social*, Jean Jacques Rousseau (1712-1778) was obsessed by the demands of life in society, by the relationships of dependence and subordination which it creates among men, and by the rivalries and enmities which it engenders. Society, which should bring men together, in fact sets them apart and makes them enemies to their fellow men. Rousseau first became aware of this paradox not in terms of clearly conceived ideas but of deeply felt experience. In the course of a tormented and eventful life, he felt poignantly the injustice of a social order founded on the inequality of status and the impossibility of achieving happiness in such a society. To an unusual extent it is necessary to know about Rousseau's life in order to understand his work.

Early life. Rousseau was born in Geneva; although he spent the greater part of his life in France, he always felt himself to be a stranger there and made a great point of indicating his attachment for his "fatherland" by adding to his name the title "Citizen of Geneva."

It was in Geneva and in the surrounding countryside that his childhood years were spent. He never knew his mother, who died shortly after his birth, and, as he often noted, the education he received from his father was an unusual one. Isaac Rousseau was a capable watchmaker but an odd and temperamental man, and he all too early inculcated in his son a taste for reading by having him read novels and Plutarch's *Lives*. "When I was six," Rousseau wrote in a letter to Malesherbes, dated January 12, 1762, "I happened upon Plutarch; at eight I knew him by heart. I had read all the available novels, which had made me shed buckets of tears before the age when the heart usually takes an interest in such works. From them was formed in me that heroic and romantic taste which up to the present time has continued to grow, and which made me disgusted with everything except what resembled my follies" (*Oeuvres complètes*, vol. 1, p. 1134 in the Gallimard edition). Alluding to this reading, as well as to his precocious sensibility and troubling sensuality, Rousseau later wrote in the *Confessions*: "My childhood was not that of a child; my feelings and thoughts were always those of a man" (*ibid.*, p. 62).

At the age of 13, Rousseau was apprenticed to an engraver, who punished him harshly for petty thefts and pranks. After three years, a chance event, in which later he was to see a sign of des-

tiny, put an end to this apprenticeship and led to his departure from his native city. One Sunday in March 1728, he returned from a country walk too late to enter the city gates, and rather than return to Geneva the next morning, he chose to try his fortune. This adventure led him eventually to Turin, to the Hospice of the Holy Ghost, where, abjuring "the religion of my fathers," he became a Roman Catholic, apparently without serious awareness of what such a conversion meant for him. "I became a Catholic," he was to say at the end of his life, "but I always remained a Christian" (*ibid.*, vol. 1, p. 1013).

Self-education. Unable to find a means of livelihood in Turin, Rousseau went in 1729 to Annecy in Savoy, to live with Mme. de Warens, "the good woman" who had received him hospitably when he first fled from Geneva. The time he spent in Savoy, from 1729 to 1742, was a decisive period for his formation as a writer.

About thirty years old, Mme. de Warens was herself a new convert to Catholicism and something of an adventuress. She encouraged Rousseau's affection for her, and he lived with her until he finally resolved to seek his fortune in Paris. At first she treated him as a son, but after a time he became her lover. Rousseau's life in Savoy was both happy and free; according to the *Confessions*, those 12 years were "the one short period of happiness of my entire life."

The years he spent with Mme. de Warens, not only at Annecy but also at Chambéry and Les Charmettes, were chiefly years of study. Whereas the other great writers of the eighteenth century, Montesquieu, Voltaire, and even Diderot, were formally educated, Rousseau taught himself. Everything he learned, as he said in the *Confessions* with all the pride of an autodidact, he learned by himself. At Les Charmettes, his passion for reading was reawakened, and he ranged over all domains of learning, carefully jotting down excerpts from his readings in notebooks that became veritable "storehouses of ideas." When he began studying the sciences, he did geometry exercises, performed chemistry experiments, and even made astronomical observations. Finally he learned music. Although "born for this art," he had received no musical instruction, and at the time of his arrival at Annecy, he could hardly read a musical score. Ten years later, he was to be able both to compose music and to direct an orchestra; by a "determined study of the obscure books of Rameau" and other scholarly works, he had acquired a profound knowledge of musical composition and theory.



It was with this baggage of knowledge that he left Savoy in 1742. Like his departure from Geneva in 1728, this was also a flight. Rousseau felt that he no longer occupied the same place in Mme. de Warens' life. But this time he left with a specific goal—that of submitting to the Academy of Sciences the system of musical notation he had invented.

Early success. Rousseau arrived in Paris during the summer of 1742, equipped with only a few letters of recommendation and his system of musical notation. He had come to the city in quest of fame, but for ten years he was obliged, as Jean Guéhenno put it, to "mark time in the crowd of men of letters" (1948-1952).

He succeeded in interesting various ladies of Parisian society in his fate, and one of them obtained employment for him as secretary to the new French ambassador in Venice. The stay in Venice was important because it was there that he first conceived the ideas that were fully developed in his *Contrat social* (1762a). His conflicts with the ambassador, however, were such that after a year he was dismissed. Naively, he expected that the ambassador would be blamed and the secretary vindicated, but he was quickly disillusioned. From then on, he was convinced that there was no reason to expect any kind of justice from a social order founded on inequality of status.

Upon his return to Paris, he began a liaison with an uneducated young domestic servant, Thérèse Levasseur, who became his lifetime companion and who served him both as housekeeper and as nurse. According to the *Confessions*, several children were born from this liaison; all of them were placed in the Foundlings' Home. Although Rousseau later felt that it had been a "mistake" to abandon his children, he nevertheless did so, according to the *Confessions*, "without the slightest scruple." He could not afford to rear a family on his earnings as a writer; besides, he was only following "local custom," since in 1745 almost a third of all the infants born in Paris were abandoned to institutional care.

Two successes, one literary, the other musical, lifted Rousseau out of obscurity. In 1749, the Academy of Dijon, which was then one of the flourishing provincial academies, announced an essay contest dealing with the following question: "Has the progress of the sciences and arts contributed to the purification of morals?" Rousseau entered the contest and won the prize by replying "No." In his *Discours sur les sciences et les arts* (1750) he first developed a theme to which he reverted

constantly in his later work: the contrast between ancient and modern peoples. He found the former to be virtuous, by reason of their ignorance, their simplicity, and even their coarseness of manner, whereas the latter are too subtly refined not to be corrupt. Yet had historians paid closer attention to the *Discours*, they would have avoided the mistake of reducing Rousseau's thought to a simple antithesis between man's natural goodness and his corruption by society. This antithesis is doubtless of great importance, but it should not obscure his more basic premise that all peoples are precisely what their social institutions make them. Rousseau's essay provoked various polemics, including a refutation by the Polish king Stanislas, and these polemics undoubtedly gave Rousseau very useful publicity.

A short time later, Rousseau the musician was to have the same success as Rousseau the writer. In 1752 *Le devin du village* was played twice at Fontainebleau for Louis xv, and the following year several times at the Opéra. "Le devin du village," Rousseau said in the *Confessions*, "added the finishing touch to my popularity, and I became the most sought after man in Paris" (*Oeuvres complètes*, vol. 1, p. 369 in the Gallimard edition). Indeed, at this period he was known primarily as a musician, but he had the wisdom to give up music in order to devote himself entirely to literature.

Early writings about society. It was in the ten years of intense literary production from 1752 to 1762 that Rousseau composed his most important writings. The first of these was the *Discours sur l'origine de l'inégalité* (1755). It was a work of circumstance, since Rousseau was again answering a question proposed by the Academy of Dijon. However, the subject was so close to his own concerns that he could not have chosen a more appropriate one himself. For Rousseau, inequality is the original evil, the one that engenders all others. Thus, by opposing "the equality which nature established among men," to "the inequality which men have instituted," Rousseau, in this second *Discours*, made a case against life in society similar to the case against civilization that he had made in the first.

His article "Économie politique," published in 1755 in the fifth volume of the *Encyclopédie*, was written in the same period as the second *Discours*. The article is a short political treatise in which the author expounded his ideas on government, public education, and state finances. While the second *Discours* is essentially a work of criticism, intended to underline all the evils created for man

by life in society, by private property, and by social inequality, "Économie politique" shows the positive, constructive side of Rousseau's thought.

Montmorency. Rousseau had completed the writing of the *Discours sur l'origine de l'inégalité* in 1754 and had it printed in Holland in 1755. He dedicated it to the Republic of Geneva, and in the dedication he praised his native city, its government, and its magistrates, giving himself up completely to his republican enthusiasm and patriotic zeal. Upon his return to his native land in June 1754, he decided to become a Protestant again in order to resume his rights as a Genevan citizen. He attached so little importance to matters of cult and to those questions which divided the Christian confessions that he failed for the second time to attribute religious significance to the change. "Believing that there were not two ways for a reasonable man to be a Christian, I believed also," he wrote in the *Confessions*, "that all matters of formality and discipline in any state fall within the purview of the laws" (*Oeuvres complètes*, vol. 1, p. 392 in the Gallimard edition). Thus, his return to the religion of his fathers was, in effect, a civic act rather than a religious one.

In the patriotic joy of his homecoming, Rousseau may have intended to settle in Geneva. If such was indeed his intention, it was an idea which he quickly abandoned. Voltaire had just moved into Les Délices, quite near Geneva, and Rousseau knew well enough that there was not room for both of them in his native city. He also knew that he would be less free in Geneva than in France to publish his writings under his own name. This consideration weighed heavily in his decision to accept the invitation of Mme. d'Épinay to move to L'Ermitage, a summer house which she had prepared for him on her estate on the edge of the Montmorency forest. After so many years spent in Paris, "city of noise, smoke, and mud," Rousseau was delighted to return to country life. However, he stayed there less than two years; he moved in during April 1756 and left in December 1757, after a memorable quarrel with Mme. d'Épinay, with Grimm, and even with Diderot. The origins of the quarrel lay in Rousseau's infatuation with Sophie d'Houdetot, a sister-in-law of Mme. d'Épinay. The young countess encouraged his passion, if she did not share it; it was the great passion of Rousseau's mature life. The novel he was writing, *La nouvelle Héloïse* (1760), was one source of his involvement: he fell in love with Mme. d'Houdetot, seeing in her his own heroine Julie—a case of literary creation giving rise to events in real life.

The amorous involvement at L'Ermitage lasted barely four months. Mme. d'Houdetot ended all communication with Rousseau as soon as she saw herself compromised by his indiscretions. Rousseau suspected Diderot of having divulged the secret and, feeling himself betrayed, decided not merely to break with Diderot but to break with him publicly, which he did in the Preface to his *Lettre à M. d'Alembert* (1758).

Despite the break with Diderot and Mme. d'Épinay, Rousseau refused either to return to Paris or to leave the Montmorency forest, which he had made his "study." He stayed on at Montmorency for more than four years, first in a house he rented in the Mont-Louis park, and then in the Petit Château of the marshal of Luxembourg.

While at L'Ermitage, Rousseau abandoned once and for all the idea of writing a work entitled "La morale sensitive ou le matérialisme du sage." He postponed the final version of his *Dictionnaire de musique* (which was not finished until 1764 in Môtiers and not published until 1768 in Paris). He tried to finish a work on the abbé de Saint-Pierre as quickly as possible, and therefore he unfortunately gave up his original plan to analyze all the abbé's *projets* and to write his biography, for which, he felt, he "had collected some rather good materials." Instead, he dealt only with the *Projet de paix perpétuelle* and the *Polysynodie* (a proposal for multiple advisory councils), presenting a "critique" after each.

Shortly after moving to Mont-Louis, in February 1758, Rousseau wrote the *Lettre à M. d'Alembert sur les spectacles*. This was again a work prompted by circumstance. D'Alembert had written the article "Genève," published in the fall of 1757 in the seventh volume of the *Encyclopédie*. Two passages in the article, both inspired by Voltaire, caught Rousseau's attention—a malicious praise of the Genevan clergy ("a number of Genevan pastors have no other religion than pure socinianism") and a plea for the establishment of a theater in the city. Preferring to let the pastors defend themselves, Rousseau passed rapidly over the religious question, and he devoted the greater part of the work to public entertainment generally, not simply to the theater, as is often supposed. Indeed, the *Lettre à M. d'Alembert* contains, in addition to Rousseau's criticism of the theater and of actors, a rough outline of a sociology of entertainment, considering entertainment in relation to the society for which it is intended. Thus, monarchies can support the luxury of a theater, but republics, and particularly small cities, require forms of enter-



tainment more appropriate to them, such as festivals, public balls, games, dances, concerts, gymnastic competitions, and military parades; in short, free entertainment held in the open air.

According to the *Confessions*, Rousseau wrote the *Lettre à M. d'Alembert* in only three weeks and then devoted his time to the completion of his novel, which was printed in Holland by Marc-Michel Rey. *La nouvelle Héloïse* appeared in the beginning of 1761, under the title *Lettres de deux amans, habitans d'une petite ville au pied des Alpes, recueillies et publiées par J.-J. Rousseau*. It enjoyed an immediate and huge success with the public and ever since has been counted among the masterpieces of European literature.

**Thoughts on childhood.** With *La nouvelle Héloïse* finished, Rousseau returned to two other major projects: *Émile* (1762b), which was to present his system of education, and the "Institutions politiques." He did finish a full-scale version of the first; the second was reduced to a short treatise, to which he gave the title *Du contrat social, ou principes du droit politique* (1762a). Both works have become classics, in pedagogy and political science, respectively.

An innovator in almost all domains, Rousseau gave proof of his originality especially in his treatise on education, which has so much in common with the pedagogy of our own time. His object in *Émile* was not to make a craftsman or a gentleman of his pupil; his aim was to "educate him to be a man," for, as he said, "a natural education should fit him for any human vocation." Instead of letting himself become imprisoned within the narrow framework of pedagogy, Rousseau conceived of his book as a reflection upon man and man's estate. "It matters little to me," he said further, "that I have written a novel. Human nature itself makes a rather fine novel" (*Oeuvres complètes*, vol. 2, p. 387 in the Hachette edition). But while the object of education is to form a man, its subject is the child. Rousseau reproached his predecessors for having forgotten this fact, and he blamed adults in general for being unable to imagine any mentality other than their own. "Childhood," he said, "has its own ways of seeing, thinking, and feeling; nothing is more foolish than to try to replace them with ours" (*ibid.*, vol. 2, p. 37). There is, then, a mentality peculiar to the child, and the first aim must be to understand it. Rousseau is probably the first to have linked the science of education to the scientific understanding of the child, and for this reason he is rightfully recognized today as the precursor and even the founder of child psychology.

Nevertheless, his conception of childhood re-

mained theoretical rather than experimental, for it rested almost entirely on an analogy drawn between the state of childhood and the state of nature. The child is as alone in human society as the savage in the thick of his woods. It is because of the child's natural solitude that Rousseau warned against premature socialization. "Parents who live in civilized society bring their children into it too young" (*ibid.*, vol. 2, p. 51). The second mistake to avoid is that of reasoning with children, since the other characteristic common to both childhood and the state of nature is a life governed by pure instinct; reason awakens and develops only with age and through social relationships. This is why Rousseau divided education into two distinct phases: the first, education concerned with the world of things, occupies the period of childhood; the second, moral education, must be put off until the beginning of adolescence, when the young man, having become involved in social relationships, develops his reason naturally, rather than prematurely, because he needs it to control his awakening passions. "The proper study of man is the study of his relationships. As long as he knows himself only through his physical being, he must study himself in relation to things; this is the program of his childhood: when he begins to sense his moral being, he must study himself in relation to men; this is the program of his entire life . . ." (*ibid.*, vol. 2, p. 184).

**The "Social Contract."** While *Émile* is a treatise on man, the *Contrat social* is a treatise on the citizen, but both are inspired by the same ideal—they are intended to safeguard liberty, which is constantly threatened by social relationships and social institutions. "The greatest good," said Rousseau in *Émile*, "is not authority, but liberty." Relationships of authority are almost always arbitrary or tend to become so. "Dependence among men being without order or rule," authority gradually transforms itself into domination and obedience into servitude. If, therefore, liberty is to be preserved, this dependence among men must be eliminated or at least subjected to regulation. From this point of view, the requirements of liberty create the same task for the legislator as for the educator. Natural education, we have seen, consists in withdrawing the child from dependence upon adults in order to subject him to dependence upon things. Similarly, the art of politics consists in making each citizen extremely dependent upon the polis in order to free him from dependence upon other citizens (*Contrat social*, book 2, chapter 12 in *Oeuvres complètes*, vol. 3, p. 394 in the Gallimard edition).

Not only are the various clauses of the *Contrat*

social finally reducible to a single stipulation—"the total surrender of each individual with all his rights to the whole community" (*ibid.*, vol. 3, p. 360)—but the contract or pact correspondingly establishes the absolute supremacy of the state over all its members. Unlike Locke or Montesquieu, Rousseau asserted no fundamental law, nor did he place constitutional limitations on sovereignty. "It is the essence of Sovereign Power," Rousseau said in 1764, in his *Lettres écrites de la montagne*, "to be illimitable; it is omnipotent or it is nothing" (*ibid.*, vol. 3, p. 826). Rousseau's political doctrine excludes any balance or equilibrium of powers; there is in the state only one supreme power, to which all others are subordinated. The sovereign people can at any time change its rulers and its laws or modify the form of its governmental administration and the constitution of the state. In principle there is nothing it cannot do.

It is understandable, therefore, that a considerable number of historians have reproached Rousseau for having paid only lip service to the cause of liberty and for having in fact opened the gates to despotism, collectivism, and totalitarianism. In France, he has been blamed for having sacrificed, through his theory of sovereignty, the rights of the individual to the omnipotence of the state. This error, said Benjamin Constant in 1815, made his *Contrat social*, "which is often invoked in behalf of liberty, the most terrible weapon of all types of despotism" ([1815] 1957, p. 1105). A century later, the jurist Léon Duguit called Rousseau the "source of all the doctrines of dictatorship and tyranny, from the doctrines of the Jacobins in 1793 to the Bolshevik doctrines of 1920" (1922, p. 135). In England, and elsewhere, the *Contrat social* did not arouse such impassioned discussion as in France. Nevertheless, it is significant that the scholarly editor of *The Political Writings of Jean-Jacques Rousseau*, C. E. Vaughan, considered it anti-Lockean: "... while the Contract of Locke," he says, "is expressly devised to preserve and confirm the rights of the individual, that of Rousseau ends, and is intended to end, in their destruction. . . . One is a charter of individualism. The other, with certain crucial qualifications to be noted shortly, is an extreme form of collectivism" ([1915] 1962, p. 48). More recently, J. L. Talmon stated that Rousseau's ideas were to blame for the "rise of totalitarian democracy" (1952).

These interpretations, based on the concepts of the "total surrender" of individual rights ("l'aliénation totale") and of the absolute sovereignty of the state over all its members, draw conclusions from the *Contrat social* that are fundamentally opposed

to the intentions of its author. Indeed, for Rousseau, liberty was the most precious of possessions, a gift which nature has made to men. They can no more be deprived of it rightfully than they can be deprived of life itself; nor can they be permitted to divest themselves of it for any price whatsoever. The social pact should not be interpreted as abrogating, in effect, a right which Rousseau declared inalienable and inseparable from the essential character of man. Rousseau affirmed that the surrender of rights is only apparent and that in the end individuals regain the rights which they appear to have given up. "Instead of a surrender," he said, "they have only made a profitable exchange" (*Contrat social*, book 2, chapter 4 in *Oeuvres complètes*, vol. 3, p. 375 in the Gallimard edition). In effect, the social contract is designed to secure or to restore to individuals in the state of civilization the equivalent of the rights they enjoyed in the state of nature.

The best interpreters of Rousseau, including Vaughan, have asserted that the theory of the general will constitutes a guarantee of individual rights, rather than their sacrifice. By the social pact individuals place themselves under "the supreme direction of the general will," and, in Rousseau's definition, it is "the exercise of the general will" that is sovereignty. Rousseau replaced the sovereignty of kings with the sovereignty of the people. But even this is not the essential safeguard to liberty. While the sovereign people cannot harm all its members—for this would be to harm itself—nonetheless this scarcely constitutes assurance that it cannot harm a single individual or that it cannot sacrifice the minority to the interests of the majority. This is why Montesquieu, for example, spoke of the "despotism of all." But such an eventuality is rendered impossible in Rousseau's system by the fact that the general will can never rule on individual cases or issues, but imposes the same sacrifices and confers the same advantages upon all citizens. If one individual is wronged, all are thereby wronged. The equality established among all citizens by the general will is their protection against any abuse of power by the sovereign. Since, in Rousseau's system, the sovereign authority cannot exceed "the limits of the general covenants," it follows that sovereignty, although absolute in principle, is always limited in practice.

Rousseau presented still another argument: that the citizen, by subjecting himself to the general will and to the laws which it prescribes, obeys no one but himself and, therefore, "remains as free as he was before." This reasoning is not entirely without difficulties, however, for in point of fact



the "general will" is not the "will of all." The general will is a corporate will, the will of the corporate people. But this body is made up of citizens, and their general will *as citizens* may well be different from the particular wills they have *as individuals*. The general will cannot finally prevail in the individual heart unless the man disappears behind the citizen and changes his "nature," so to speak, in order to identify himself with the polis. To make a man a citizen, it is thus indispensable to transform the human mentality. In the *Contrat social*, Rousseau entrusted this task to the legislator, but according to *Émile*, it is the proper goal of public education, without which there can be no citizens.

Although Rousseau affirmed in Book 4 of the *Contrat social* that "since man is born free and his own master, no one can govern him without his consent" (*Oeuvres complètes*, vol. 3, p. 440 in the Gallimard edition), it is doubtful whether the "consent" of the governed in his system has the same importance as in Locke's *Two Treatises of Government*: the ideal of "government by consent" remains closer to liberal theories of the state than to the equalitarian conception of Rousseau.

By the logic of Rousseau's doctrine, it is precisely because the general will maintains equality among citizens that its exercise is not susceptible to abuse. Abuses of power and risks of oppression can come only from the intermediate body which is responsible for the enforcement of the laws or the administration of the state. (According to the terminology created by Rousseau, which subsequently became common, this intermediate body is called the government, as distinguished from the sovereign.) Sovereignty being nothing else than the exercise of the general will, the sovereign can act only by means of laws and consequently must entrust the executive power to a subordinate body. But this body, like all collective bodies, tends to place its own corporate interest before the common interest and to substitute itself for the sovereign. The true political problem for Rousseau consisted, thus, in keeping the government within the limits of its power and in preventing it from becoming master of the state by usurping that sovereignty, or legislative power, which can belong only to the people.

The great distance that separates Rousseau's political doctrine from Montesquieu's emerges clearly. Not only is Rousseau's doctrine equalitarian and antihierarchical; it is at the same time republican and antimonarchist. To be sure, he retained monarchy among the forms of government, but he saw kings simply as magistrates or officers respon-

sible for the enforcement of the laws. This qualification in effect withdraws from kings the rights of sovereignty and confers these rights solely upon the people. As Rousseau's contemporaries observed, and as the Jesuit G. F. Berthier in particular noted in his *Observations sur le Contrat social* (1789), it was tantamount to granting recognition neither to monarchy nor to aristocracy, but only to democracy.

Last years. Rousseau was clearly running great risks by publishing such a book as the *Contrat social* under his own name. It was published in Amsterdam by Marc-Michel Rey and distributed in the spring of 1762. Its circulation was forbidden in France, but no further measures were taken against it. *Émile*, however, unleashed a storm. It had been printed in Paris by Duchesne (under the fictitious editorship of Jean Néaulme of Amsterdam) and was offered for sale in May 1762. Rousseau had inserted in the work a section which had originally been separate and in which he presented his religious ideas, *Profession de foi du vicaire savoyard*. This was the pretext seized upon by the Parlement of Paris to condemn the work and to issue a writ for Rousseau's arrest on June 9, 1762. (The Parlement proceeded against him with such severity because it was about to expel the Society of Jesus from France and wished to strike at unbelievers at the same time.) However, the writ of arrest was not enforced with much haste, and Rousseau was able to leave France without being harassed.

Where was he to go? Even had he wished, he could not have gone to Geneva, for on June 19, ten days after the writ of the Parlement, the Genevan Advisory Council condemned both *Émile* and the *Contrat social*, charging that both works were "destructive of the Christian religion and of all governments." During his years of exile Rousseau constantly sought a suitable place to stay, an agonizing problem for a man 50 years old and afflicted with a painful infirmity. Switzerland was the country he preferred, but where in Switzerland? First Rousseau resided for three years—July 1762–September 1765—at Môtiers in the principality of Neuchâtel, which then belonged to the king of Prussia. There he prepared two polemical writings in defense of his work, an open letter addressed to the archbishop of Paris, Christophe de Beaumont, which he composed during the winter of 1762 and published the following year, and the *Lettres écrites de la montagne* (1764), an answer to the *Lettres écrites de la campagne*, written by the prosecuting attorney, J.-R. Tronchin. He declined responsibility, however, for the agitation on his behalf by his Genevan

friends and preferred to renounce his Genevan citizenship in 1763, rather than involve himself in a political quarrel. His vocation was to write, not to become a party leader.

Rousseau was driven from Môtiers by a popular demonstration, the importance of which he exaggerated, and subsequently he was expelled from the Île Saint-Pierre by the Bernese authorities. He returned to Paris without being harassed, but only for a temporary stay. Then, at the invitation of David Hume, he agreed to go to Scotland, but quarreled with his host. Rousseau stayed almost a year and a half in England, first on the outskirts of London, then at Wootton in Derbyshire, and finally returned to France in May 1767. Here, for three years, he lived the life of a hunted man, first at Trye-le-Château in the Oise region, then at Bourgoin and at Monquin in the province of Dauphiné. During this time he changed not only his residence but also his name, signing his letters as "Renou."

In 1770 Rousseau received assurance that he could reside in Paris without being disturbed and thereupon moved into small quarters on the rue Plâtrière. For several years he was obsessed to the point of madness by the idea that there was a plot or a kind of universal conspiracy against his person and his writings. He never did wholly shake off this obsession, although toward the end of his life he succeeded in regaining a certain peace of mind. In Paris he spent his time copying music (which was his livelihood), gathering herbs, and writing.

He did not stop writing after the publication of his three great works. During his stay in Môtiers in 1765, he drafted the *Projet de constitution pour la Corse*, which was never finished, and in Paris in 1772, *Considérations sur le gouvernement de Pologne*. In the first of these two works, the author of the *Contrat social* applied his "principles" to the problems of legislating for a newly independent people and in the second he proposed reforms to a proud nation threatened by powerful neighbors.

His primary concern was a vast project which he had been contemplating for a long time, his autobiography. It was an undertaking which would require several years of unceasing work. Rousseau knew this and, despairing of bringing the project to completion, wrote in 1762 the four autobiographical *Lettres à M. de Malesherbes*. Later, however, he succeeded in drafting his *Confessions*, which were composed in two different periods, from 1765 to 1767 and in the winter of 1769/1770. The *Confessions* were written largely for the pleasure of knowing himself and reliving his past, but the

three dialogues, written in the period 1772-1775 and entitled *Rousseau, juge de Jean-Jacques*, were composed while he was haunted by fears of persecution. He wrote this strange book to vindicate himself: "I should willingly consent," he said, "to not existing at all in the memory of men, but I cannot consent to remain there forever dishonored by calumny" (*Oeuvres complètes*, vol. 1, p. 953 in the Gallimard edition). His obsession with a conspiracy, which dominated the dialogues, persisted to some extent in his *Rêveries du promeneur solitaire* (1782), but Rousseau had largely regained his serenity when he wrote these reflections. None of his autobiographical writings was published until after his death, which came suddenly on July 2, 1778, in Ermenonville. He was buried at Ermenonville, and pilgrimages were made to his tomb there until the Convention ordered his remains transferred to the Panthéon.

In our day, sociologists and ethnologists have recognized Rousseau, the self-styled "apologist of nature," as a forerunner of the social sciences and perhaps, even, as their founder. Émile Durkheim said succinctly of the *Discours sur l'origine de l'inégalité*: "Rousseau demonstrated a long time ago that if all that comes to man from society were peeled off, there would remain nothing but a creature reduced to sense experience and more or less undifferentiated from the animal" (1906, p. 132). To rise above the animal stage, men must become sociable and therefore must relinquish the state of nature. Such appears to be the teaching of the second *Discours*. According to Rousseau, what distinguishes man from the animal is his perfectibility. But perfectibility is closely connected with sociability, which permits it to develop, whereas in the state of nature it is only a "potential faculty." Rousseau held that "Man in a state of isolation always remains the same; only in society does he progress" (*Oeuvres complètes*, vol. 3, p. 533 in the Gallimard edition). So long as he lives in the isolation of the state of nature, he exists only as a physical being. As a moral being he is formed in society, for his passions, his reason, and even his conscience become active and develop only in social relationships. It is true that perfectibility brings in its wake a multitude of evils from which man in the state of nature was exempt, but "nature does not retrogress." From this it follows that human history begins with life in society.

Such considerations as these anticipated later developments in social science. Claude Lévi-Strauss has recently shown that ethnology is indebted to



Rousseau and that it has borrowed from him the methodological rule he stated in the *Essai sur l'origine des langues*: "To study men, we must look close by; to study man, we must learn to look afar; if we are to discover essential characteristics, we must first observe differences" (*Oeuvres complètes*, vol. 1, p. 384 in the Hachette edition). This passage shows clearly what kind of research Rousseau thought necessary for the understanding of man; sociology had its objective fixed in advance, so to speak, by the precept of *Émile*: "We must study society by studying men, and men by studying society. Those who expect to treat political and moral philosophy as separate entities will never understand anything about either" (*ibid.*, vol. 2, p. 206).

ROBERT DERATHÉ

[See also CONSENSUS; CONSTITUTIONS AND CONSTITUTIONALISM; GENERAL WILL; MAJORITY RULE; NATURAL LAW; SOCIAL CONTRACT; SOVEREIGNTY; STATE; UTOPIANISM, article on UTOPIAS and UTOPIANISM. Other relevant material may be found in POLITICAL THEORY and in the biography of DURKHEIM.]

#### WORKS BY ROUSSEAU

- (1750) 1946 *Discours sur les sciences et les arts*. Critical edition. The Modern Language Association of America, Monograph Series, No. 15. New York: The Association; London: Oxford Univ. Press. → First published as *Discours qui a remporté le prix à l'Académie de Dijon*.
- (1750-1755) 1964 *The First and Second Discourses*. Edited with an introduction and notes by Roger D. Masters. New York: St. Martin's. → Contains *Discourse on the Sciences and Arts* and *Discourse on the Origin and Foundation of Inequality*. First published in French.
- (1752) 1964 *Le devin du village*. Volume 2, pages 1093-1114 in Jean-Jacques Rousseau, *Oeuvres complètes*. Paris: Gallimard.
- (1755) 1941 *Discours sur l'origine et les fondements de l'inégalité parmi les hommes*. Cambridge Univ. Press.
- (1758) 1948 *Lettre à M. d'Alembert sur les spectacles*. Critical edition, by M. Fuchs. Lille: Glard. → Translated into English as *Politics and the Arts: Letter to M. d'Alembert on the Theatre* and published by the Free Press in 1960.
- (1761) 1964 *La nouvelle Héloïse*. Volume 2, pages 5-796 in Jean-Jacques Rousseau, *Oeuvres complètes*. Paris: Gallimard.
- (1762a) 1961 *The Social Contract*. London: Dent. → First published in French.
- (1762b) 1963 *Émile*. London: Dent; New York: Dutton. → First published in French.
- (1762c) 1873 *Profession de foi du vicaire savoyard*. Volume 2, pages 236-327 in Jean-Jacques Rousseau, *Oeuvres complètes*. Paris: Hachette.
- (1762d) 1928 *Lettres à M. de Malesherbes*. London: Scholartis Press.
- 1763 *Jean-Jacques Rousseau, citoyen de Genève, à Christophe de Beaumont, archevêque de Paris*. Amsterdam: Rey.

- (1764) 1962 *Lettres écrites de la montagne*. Neuchâtel (Switzerland): Ides et Calendes.
- (1765-1770) 1960 *The Confessions of Jean Jacques Rousseau*. 2 vols. New York: Dutton; London: Dent. → Written between 1765 and 1770; first published between 1782 and 1789 in French.
- (1772) 1873 *Considérations sur le gouvernement de Pologne*. Volume 5, pages 237-302 in Jean-Jacques Rousseau, *Oeuvres complètes*. Paris: Hachette.
- (1772-1775) 1959 *Rousseau juge de Jean-Jacques*. Volume 1, pages 657-992 in Jean-Jacques Rousseau, *Oeuvres complètes*. Paris: Gallimard.
- (1781) 1877 *Essai sur l'origine des langues*. Volume 1, pages 370-408 in Jean-Jacques Rousseau, *Oeuvres complètes*. Paris: Hachette.
- (1782) 1948 *Les rêveries du promeneur solitaire*. Critical edition by John S. Spink. Paris: Didier. → An English translation, *The Reveries of a Solitary*, was published by Routledge in 1927.

*Correspondance complète*. Édition critique, établie et annotée par R. A. Leigh. Vols. 1-4. Geneva: Institut et Musée Voltaire, Les Délices, 1965- → Volume 1 contains letters 1-97, dating from 1730 to 1744.

*Correspondance générale de J.-J. Rousseau*. 20 vols. Annotated by Théophile Dufour. Paris: Colin, 1924-1934. → Letters dated 1728-1778.

*Oeuvres complètes*. 12 vols. Paris: Hachette, 1871-1877. → Customarily designated the "vulgate" edition of Rousseau's works. See especially Volumes 1 and 2: *Discours* and *Émile*. Translations have been provided by the editors.

*Oeuvres complètes*. 3 vols. Edited by Bernard Gagnebin and Marcel Raymond. Bibliothèque de la Pléiade. Paris: Gallimard, 1959-1964. → Volumes published as of 1964 are the following: Volume 1: *Les confessions*; *Autres textes autobiographiques*. Volume 2: *La nouvelle Héloïse*; *Théâtre*; *Poésies*; *Essais littéraires*. Volume 3: *Du contrat social*; *Écrits politiques*. Translations have been provided by the editors.

*Political Writings*. 2 vols. Edited from the original manuscripts and authentic editions, with an introduction and notes by C. E. Vaughan. New York: Wiley, 1962. → This edition was first published in 1915.

#### SUPPLEMENTARY BIBLIOGRAPHY

- BERTHIER, GUILLAUME F. 1789 *Observations sur le Contrat social de J.-J. Rousseau*. Paris: Mériogot le jeune.
- BURGELIN, PIERRE 1952 *La philosophie de l'existence de J.-J. Rousseau*. Paris: Presses Universitaires de France.
- CASSIRER, ERNST (1932) 1954 *The Question of Jean-Jacques Rousseau*. Translated and edited by Peter Gay. New York: Columbia Univ. Press. → First published in German. A paperback edition was published in 1963 by Indiana University Press.
- CHAPMAN, JOHN W. 1956 *Rousseau: Totalitarian or Liberal?* New York: Columbia Univ. Press.
- COBBAN, ALFRED (1934) 1964 *Rousseau and the Modern State*. 2d ed. London: Allen & Unwin.
- CONSTANT DE REBECQUE, HENRI BENJAMIN (1815) 1957 *Principes de politique*. Pages 1097-1249 in Henri Benjamin Constant de Rebecque, *Oeuvres*. Bibliothèque de la Pléiade, No. 123. Paris: Gallimard.
- COURTOIS, LOUIS J. 1923 *Chronologie critique de la vie et des oeuvres de Jean-Jacques Rousseau*. Société Jean-Jacques Rousseau, *Annales* 15:1-364.
- DERATHÉ, ROBERT 1948 *Le rationalisme de J.-J. Rousseau*. Paris: Presses Universitaires de France.

- DERATHÉ, ROBERT 1950 *Jean-Jacques Rousseau et la science politique de son temps*. Paris: Presses Universitaires de France.
- DUGUIT, LÉON 1922 *Souveraineté et liberté*. Paris: Alcan.
- DURKHEIM, ÉMILE 1906 Détermination du fait moral. *Société Française de Philosophie, Bulletin* 6:113-138. → Letters and discussions of Durkheim's article appear on pages 169-212. The article and discussions were partially translated and published in 1953 in Durkheim's *Sociology and Philosophy*.
- DURKHEIM, ÉMILE (1918) 1960 Rousseau's *Social Contract*. Pages 65-138 in Émile Durkheim, *Montesquieu and Rousseau: Forerunners of Sociology*. Ann Arbor: Univ. of Michigan. → First published in French.
- FETSCHER, IRING 1960 *Rousseaus politische Philosophie*. Neuwied (Germany): Luchterhand.
- GUÉHENNO, JEAN (1948-1952) 1966 *Jean-Jacques Rousseau*. New ed. 2 vols. New York: Columbia Univ. Press. → Volume 1: 1712-1758. Volume 2: 1758-1788. First published in French. An excellent biography.
- HENDEL, CHARLES W. (1934) 1962 *Jean-Jacques Rousseau, Moralist*. Indianapolis, Ind.: Bobbs-Merrill.
- HUBERT, RENÉ 1928 Rousseau et l'Encyclopédie: *Essai sur la formation des idées politiques de Rousseau (1742-1756)*. Paris: Gamber.
- Jean-Jacques Rousseau*, by Samuel Baud-Bovy et al. 1962 Neuchâtel (Switzerland): La Baconnière.
- Jean-Jacques Rousseau et son œuvre: Problèmes et recherches*. 1964 Paris: Klincksieck. → Commemorative work published on the occasion of the 250th anniversary of the birth of Rousseau. Colloquy of the Collège de France, October 16-20, 1962.
- LACHARRIÈRE, RENÉ DE 1963 *Études sur la théorie démocratique: Spinoza-Rousseau-Hegel-Marx*. Paris: Payot.
- MASSON, PIERRE M. 1916 *La religion de J.-J. Rousseau*. 3 vols. Paris: Hachette.
- PLAN, PIERRE P. 1953 *Table de la Correspondance générale de J.-J. Rousseau*. With an introduction by Bernard Gagnebin and some unpublished letters by Rousseau. Geneva: Droz.
- RANG, MARTIN 1959 *Rousseaus Lehre vom Menschen*. Göttingen: Vandenhoeck & Ruprecht.
- RAVIER, ANDRÉ 1941 *L'éducation de l'homme nouveau: Essai historique et critique sur le livre de l'Émile de J.-J. Rousseau*. 2 vols. Lyon: Riou.
- RAYMOND, MARCEL 1962 *Jean-Jacques Rousseau: La quête de soi et la rêverie*. Paris: Corti.
- Rousseau et la philosophie politique*. *Annales de philosophie politique*, No. 5. 1965 Paris: Presses Universitaires de France.
- SÉNELDER, JEAN 1950 *Bibliographie générale des œuvres de J.-J. Rousseau*. Paris: Presses Universitaires de France.
- SOCIÉTÉ JEAN-JACQUES ROUSSEAU, GENEVA *Annales*. → Published since 1905. Contains original studies, bibliographical news, and reviews.
- STAROBINSKI, JEAN 1957 *Jean-Jacques Rousseau: La transparence et l'obstacle*. Paris: Plon.
- TALMON, JACOB L. (1952) 1965 *The Origins of Totalitarian Democracy*. 2d ed. New York: Praeger.
- VAUGHAN, C. E. (1915) 1962 Introduction. Volume 1, pages 1-117 in Jean-Jacques Rousseau, *Political Writings*. New York: Wiley.
- VOSSLER, OTTO 1963 *Rousseaus Freiheitslehre*. Göttingen: Vandenhoeck & Ruprecht.

## RULE OF LAW

See CONSTITUTIONAL LAW; JUDICIAL PROCESS; JUDICIARY.

## RULES OF THE GAME

The concept "rules of the game" is frequently used in contemporary political science, albeit with a variety of meanings. Most commonly, the concept denotes widely shared beliefs about *how* the government, or various categories of political actors, *ought* to behave. "Rules of the game" are, therefore, essentially normative and procedural. The concept is rarely used to refer to formal or written rules—statutes, constitutions, court decisions, and the like—but usually refers to informal or unwritten rules, attitudes, and expectations. There is far less agreement, however, on the precise content of these rules, on how widely shared they are, or on what functions they perform.

The historical origins of the concept are as obscure as its psychological appeal is clear. Politics—particularly as practiced in regimes recognizing the legitimacy of overt, organized opposition and successfully limiting the use of force and fraud as political techniques—has many similarities to a game. Competing players pursue conflicting goals by means of varying strategies within the context of an agreed-upon set of rules. The outcome of the game—be it politics, football, or poker—depends partly upon the relative skill of the players and partly on chance factors over which the players have little or no control. These similarities between politics and games occurred to politicians and political journalists generations ago (Kent 1923), well before political scientists made use of the metaphor or von Neumann and Morgenstern (1944) published their mathematical theory of games [see GAME THEORY]. Recently, political scientists have sought to determine the content and consequences of the rules of the game of politics through empirical observation, rather than deriving them deductively from an abstract model of a generalized game, as in mathematical game theory. The latter approach results in rules about strategies that rational players will follow in various hypothetical conflict situations; the "rules of the game" that political scientists have written about are widespread normative beliefs which exist in the real world of politics.

### Rules of the game and political democracy

Political philosophers of varying persuasions—Edmund Burke, Walter Bagehot, Alexis de Tocque-



ville, Harold Laski, to name a few—have argued that political democracy “. . . presupposes a people so fundamentally at one that they can safely afford to bicker” (Lord Balfour quoted in Friedrich [1942] 1950, p. 153). It is characteristic of this body of literature that no attempt is made to specify what these “fundamentals” are. If this line of argument implies that there can be no basic cleavages within a democratic polity, or that a democratic citizenry must be in agreement on the goals of governmental action, this position is logically absurd and historically false (*ibid.*, chapter 5). But it is possible to interpret at least some of these writers as meaning only that the mass of the citizenry in a political system must be agreed on the rules of the game before political democracy is possible. This more restricted view lies at the heart of a recent, empirically oriented political theory. [See DEMOCRACY.]

David B. Truman (1951) assigns an important role to mass acceptance of “the rules of the game” in the functioning of American democracy. To Truman and his followers, the interest group—a set of persons interacting on the basis of one or more shared attitudes—is the basic unit of political analysis. Politics is visualized as the patterned interactions of these groups. A number of factors, over and above simple competition between interest groups, result in the stability of such a system. First, the average person belongs to a number of interest groups. In order to grow in size and to maintain internal cohesion, an interest group must, therefore, consider the other interests and group affiliations of its members. Second, interest groups vary in the extent of their formal organization and their degree of involvement in politics. Every shared attitude (interest) can serve as the base of a group; every group not currently making claims on others through attempts to influence the governmental apparatus can, if threatened, become a political interest group. Existing political interest groups must therefore pay heed to potential interest groups or face the consequences of creating their own opposition. Third—and most important for our purposes—widespread agreement on the rules of the game exists among the citizenry and officialdom. While these interests (attitudes) are not generally organized into political interest groups, they do represent powerful potential groups. Active groups, by anticipating the possible development of these attitudes into groups active in the political arena, tend to play the game of politics within the confines of these broad rules of conduct. “It is thus multiple memberships in potential groups based on widely held and accepted interests that serve as a balance wheel in a going political system like that of the

United States. . . . Without the notion of multiple memberships in potential groups it is literally impossible to account for the existence of a viable polity such as that in the United States or to develop a coherent conception of the political process” (Truman [1951] 1965, p. 514). These “widely held and accepted interests” are explicitly identified by Truman as the rules of the game. Potential groups are their guardians. [See POLITICAL GROUP ANALYSIS.]

Recent research on the attitudes of the electorate in the United States casts doubt upon this line of analysis by demonstrating that, at least in the United States, mass agreement with “the rules of the game” is much less widespread than had previously been assumed to be the case. Prothro and Grigg (1960) confronted random samples of registered voters in two widely separated cities in the United States with a set of general statements about political democracy and a second series of statements which apply these general rules to specific cases. They found overwhelming consensus in favor of the general statements, but “when these broad principles are translated into more specific propositions . . . consensus breaks down completely” (p. 286). On 40 per cent of the specific statements in one city, and 60 per cent in another, a clear majority of the voters expressed antidemocratic preferences. Prothro and Grigg did find, however, overwhelming agreement with *both* the general and specific statements about democracy among the highly educated—a finding which led them to speculate that widespread agreement on the rules of the game at the elite level may be a precondition for democratic politics, but that on the mass level it apparently is not. The relative apathy of the mass electorate, and widespread habits of democratic behavior, may account for the survival of political democracy without mass consensus on the rules of the game. [See CONSENSUS.]

Robert Dahl, in his study of New Haven, Connecticut (1961), found much the same phenomenon. The same minority of the population that possesses the highest levels of political interest, skills, and resources also possesses the greatest commitment to the rules of the game of democratic politics. The vast majority of the citizenry may not agree with these rules, but lack the interest, skills, and resources to do much about it.

Herbert McClosky (1964) provides data on attitudes toward the rules of the game of a national sample of adult American citizens (Table 1). It is clear that a sizable minority of American citizens do not accept the rules of the game, at least as measured by these statements. The same question-

Table 1 — Agreement with the rules of the game in the United States

Statement	Political influentials (N = 3,020)	General electorate (N = 1,484)
	Per cent disagreeing, i.e., giving response consistent with rules of the game*	
(1) There are times when it almost seems better for the people to take the law into their own hands rather than wait for the machinery of government to act.	87	73
(2) The majority has the right to abolish minorities if it wants to.	93	71
(3) We might as well make up our minds that in order to make the world better, a lot of innocent people will have to suffer.	73	58
(4) If congressional committees stuck strictly to the rules and gave every witness his rights, they would never succeed in exposing the many dangerous subversives they have turned up.	75	53
(5) I don't mind a politician's methods if he manages to get the right things done.	74	58
(6) Almost any unfairness or brutality may have to be justified when some great purpose is being carried out.	87	67
(7) Politicians have to cut a few corners if they are going to get anywhere.	71	57
(8) People ought to be allowed to vote even if they can't do so intelligently.	66	48
(9) Bringing about great changes for the benefit of mankind often requires cruelty and even ruthlessness.	81	69
(10) Very few politicians have clean records, so why get excited about the mudslinging that sometimes goes on?	85	62
(11) It is all right to get around the law if you don't actually break it.	79	70
(12) The true American way of life is disappearing so fast that we may have to use force to save it.	87	65

\* In the original study, agreement with statements 1-7 and 9-12 and disagreement with statement 8 were taken to indicate rejection of the rules of the game. In order to avoid confusion, the percentages in the table show the proportion of the two samples indicating agreement with the rules of the game. "Don't know" responses, which averaged less than 1 per cent, were ignored.

Source: Adapted from McClosky 1964, p. 365, table 1.

naire was administered to the delegates and alternates to the Democratic and Republican national party conventions in 1956. Their responses (summarized under the "Political influentials" column in Table 1) indicate far wider agreement with the rules of the game among these political activists than among rank-and-file Americans.

Assuming that the United States is a reasonable approximation of a political democracy, these studies suggest that agreement about the rules of the game on the mass level is less important than either the philosophers or the group theorists argue. Elite agreement on these rules may well be a necessary condition for a successful political democracy, but only additional research conducted in a variety of political systems will tell us whether the hypothesized relationship between rules of the game and political democracy needs merely to be revised or whether it must be abandoned altogether.

### Rules of the game in legislatures

A second area of political science in which the concept "rules of the game" is of central importance is the field of legislative behavior research. Contemporary students have not only found that unwritten rules of behavior exist in a variety of legislative settings, but they have also been able to describe the content of these normative expectations, estimate how widely they are accepted, describe how they are enforced, and suggest their consequences for the operation of legislative bodies.

While observers of legislative bodies have long commented upon the apparent importance of informal understandings and unwritten rules in their operation, the first systematic study of the rules of the game in a legislative body was Matthews' study (1960) of the behavior of United States senators.

The first unwritten rule of Senate behavior, according to Matthews, is that new members are expected to serve a proper apprenticeship. The freshman senator receives the committee assignments the others senators do not want; the same is true of his office suite and his seat in the chamber. In committee rooms he is assigned to the end of the table. He is expected to do more than his share of the thankless and boring tasks of the Senate, to keep his mouth shut, to listen and learn. According to the unwritten rules of the chamber, the freshman is to accept such treatment as a matter of course. Those who do not, encounter thinly veiled hostility and loss of esteem.

The great bulk of the Senate's work is highly detailed, dull, and politically unrewarding. According to the rules of the game in the Senate, it is to these tasks that a senator ought to devote a major share of his time, energy, and thought. Moreover, a senator ought to specialize, to focus his time and attention on the relatively few matters which come before his committees or that directly and immediately affect his state. Still another rule of Senate behavior is that political disagreement should not



influence personal feelings, that senators should avoid personal attacks on colleagues and strive for impersonality in debate. Moreover, senators must be willing to bargain—to operate on the basis of the reciprocal trading of favors. This requires self-restraint, tolerance, and an essentially nonideological approach to political life. Finally, senators are expected to display a high degree of institutional patriotism, to believe that the United States Senate is the greatest legislative body in the world, and to prefer serving there to holding any other public office in the land.

These norms perform important functions. They provide motivation for the performance of legislative duties that would not otherwise be performed. They encourage the development of expertise and division of labor and discourage those who would challenge it. They soften the inevitable personal conflict of a legislative body so that adversaries and competitors can cooperate. They encourage senators to become "compromisers" and "bargainers" and to use their substantial powers with caution and restraint (Matthews 1960, pp. 102–103). Without these rules of the game, the Senate could hardly operate in anything like its present form.

Yet while the consensus about these norms is impressive, they are not universally accepted or lived up to. Men who come to the Senate after distinguished careers in public or private life find serving an apprenticeship a trying experience. Members of the chamber with active presidential ambitions are unlikely to build a national reputation through faithful adherence to the rules of the game. Men with highly insecure seats, or those representing unusually diverse and complex constituencies, often find the injunctions about legislative specialization and self-restraint in floor debate difficult to reconcile with the exigencies of political survival. Finally, while senators are necessarily tolerant of differences of political viewpoint, the ease with which senators can conform to these informal expectations depends upon their ideological position. A man elected to the Senate as a "liberal" or "progressive" or "reformer" is under pressure to produce legislative results. The people who voted for him want national legislative policy changed. Yet if the liberal senator gives in to the pressure for conformity to the rules of the game, he must postpone the achievement of these objectives. If he presses for these objectives regardless of his junior position, he will become tabbed a nonconformist and will lose popularity with his colleagues. This might not be so serious a sanction if respect by one's colleagues were not extremely helpful in get-

ting legislation passed. "In the Senate," one member is quoted as saying, "if you don't conform, you don't get many favors for your state. You are never told that, but you soon learn" (Matthews 1960, p. 114).

Thus, one can describe some of the rules of the game in the Senate, and suggest how they are enforced and how they contribute to the operation of that body. But there is as yet very little evidence on the extent of agreement with these norms. Ralph Huitt (1961), for example, has argued that the degree of consensus on these norms is less than Matthews' study implies.

A study of the legislatures of California, New Jersey, Ohio, and Tennessee conducted by John Wahlke, Heinz Eulau, William Buchanan, and Roy Ferguson (see Wahlke et al. 1962) is based upon interviews with virtually every member of the bicameral legislatures of these four states and is therefore able to handle this particular question more satisfactorily. The rules of the game most frequently mentioned by these legislators are presented in Table 2. The first thing to notice about this listing is its substantial similarity to the rules Matthews found to be operative in the U.S. Senate. Moreover, the findings of Wahlke and his collaborators support the view that these rules are generally understood and accepted by virtually all members of these four state legislatures. "Of all respondents, only two perceived no rules of the game; over half readily named at least four rules . . ." (Wahlke et al. 1962, p. 143). Moreover, the rules mentioned comprised, on the whole, "a body of mutually compatible rules which does not normally confront legislators with the dilemma of choosing among contradictory rules . . . not a single respondent suggested the existence of conflicting sets of rules, each with its own set of proponents . . ." (*ibid.*, p. 148). Almost all the state legislators interviewed recognized that these rules were more or less systematically enforced by such informal but highly effective sanctions as obstruction of the nonconforming members' bills, social ostracism, mistrust, and loss of political perquisites and rewards. The authors conclude that these rules of the game are highly functional for the legislative chambers they have studied. They create group cohesion and morale, promote predictability of legislator behavior, restrain and channel conflict, and expedite the conduct of legislative business.

It is at present impossible to say to what extent these unwritten rules of the game are generally found in democratic legislative bodies or are merely a product of uniquely American factors. Wahlke and his collaborators found some intriguing differ-

Table 2 — Rules of the game most frequently perceived by legislators in four American states (in per cent)<sup>a</sup>

Rules <sup>b</sup>	California (N = 104)	New Jersey (N = 78)	Ohio (N = 160)	Tennessee (N = 119)
(1) Performance of obligations: Keep your word; abide by commitments	64	47	28	24
(2) Respect for other members' legislative rights: Support other members' local bills if they don't affect you or your district; don't railroad bills through; etc.	32	26	24	47
(3) Impersonality: Don't deal in personalities; don't engage in personal attacks on other members; etc.	30	27	32	31
(4) Self-restraint in debate: Don't talk too much; don't speak about subjects on which you're uninformed	17	9	18	59
(5) Courtesy: Observe common courtesies; be friendly and courteous even if you disagree	19	19	24	26
(6) Openness of aims: Be frank and honest in explaining bills; don't conceal real purpose of bills or amendments	24	8	22	12
(7) Modesty: Don't be a prima donna or publicity seeker	9	19	23	21
(8) Integrity: Be honest, a man of integrity	13	19	18	11
(9) Independence of judgment: Don't be subservient to groups or individuals outside of legislature	16	19	11	14
(10) Personal virtue: Exhibit high moral conduct, no drunkenness or immorality	13	0	24	8
(11) Decisiveness: Take a stand, don't vacillate	10	8	11	15
(12) Unselfish service: Don't be an opportunist; don't use your legislative position for your personal advantage	5	19	14	4

a. Percentages do not equal 100, because most respondents named more than one rule.

b. Fewer than half the rules mentioned are summarized above.

Source: Adapted from Wahlke et al. 1962, pp. 146-147, table 7.1.

ences in the perceived rules of the game in the four states they studied. If the informal rules of the game serve to complement the formal aspects of legislative organization and procedures in promoting the viability of legislatures and the attainment of their goals, then they should be expected to vary—at least to some degree—as these factors change from one political system to the next.

Allan Kornberg's analysis of the rules of the game in the Canadian House of Commons (1964) is suggestive on this question. Using essentially the same techniques as Wahlke and his collaborators, he found a set of unwritten rules that were remarkably similar to those found earlier in the American Senate and the four American state legislatures. Nonetheless, there were differences. Most of these seem explicable in terms of differences in the formal structure and role of the Canadian legislature. Greater emphasis was placed in the Canadian House on channeling controversy into conflict between the parties, and less on rules that expedite the flow of legislation and stress the importance of seniority and apprenticeship. A somewhat larger number of Canadian legislators were unaware of the rules and the sanctions available to enforce them. These differences seem to reflect the fact that the Canadian House is part of a parliamentary system in which party discipline is greater than in the United States; a system where the cabinet is able to control the flow of legislative business; and a system in which seniority is of relatively little importance. The greater ignorance of Canadian members about the rules of the game and

their enforcement may reflect the fact that individual members are less free to define their own jobs in such a setting and play a less significant role, as individuals, in the legislative process. There is thus less need in such a system for effective controls over the behavior of individual legislators by their legislative peers.

Unfortunately, systematic research on the rules of the game in legislatures in other political systems is almost nonexistent. Nathan Leites (1958) has published a valuable study of patterns of legislative strategies in the French Fourth Republic, but it is not directly and explicitly concerned with the normative expectations of legislators and their functions. Gerhard Loewenberg's study of the West German Bundestag (1967) suggests that informal, unwritten rules of the game have developed which have made a formal structure, largely inherited from the Weimar Republic, viable under drastically altered circumstances. But none of these studies permits detailed comparison with the studies conducted in the United States and Canada. Many more cross-nation comparisons need to be made.

A second area of needed research on the rules of the game in legislatures deserves brief mention. Attention so far has been directed primarily at the rules of the game for legislative bodies as a whole. Yet subgroups of legislatures—committees, parties, work groups—sometimes possess relative autonomy and specialized sets of norms which are not identical with those for the whole membership. Richard Fenno's studies of the Appropriations Committee



(1962) and of the Education and Labor Committee (1963) in the U.S. House of Representatives indicate that the payoffs from comparative studies of such legislative subgroups are likely to be large. [See LEGISLATION, especially the article on LEGISLATIVE BEHAVIOR.]

The concept of the rules of the game has, therefore, demonstrated some utility in the empirical analysis of politics. But it is also evident that it overlaps other concepts used in contemporary political science: political culture, norms, ideology, values, strategies, role, and so on. Some of these competing concepts are usually more sharply defined, and in the long run, it therefore seems quite possible that the concept may pass from common use and the game metaphor may be abandoned to the mathematical game theorists. But that time is unlikely to arrive in the near future. In the meantime, the concept serves as a constant reminder of the importance of informal, normative expectations and unwritten rules in shaping political behavior.

DONALD R. MATTHEWS

[See also IDEOLOGY; NORMS; POLITICAL CULTURE. Other relevant material may be found in POLITICAL BEHAVIOR.]

#### BIBLIOGRAPHY

- DAHL, ROBERT A. (1961) 1966 *Who Governs? Democracy and Power in an American City*. New Haven: Yale Univ. Press.
- FENNO, RICHARD F. 1962 The House Appropriations Committee as a Political System: The Problem of Integration. *American Political Science Review* 56: 310-324.
- FENNO, RICHARD F. 1963 The House of Representatives and Federal Aid to Education. Pages 195-235 in Robert L. Peabody and Nelson W. Polsby (editors), *New Perspectives on the House of Representatives*. Chicago: Rand McNally.
- FRIEDRICH, CARL J. (1942) 1950 *The New Image of the Common Man*. Enl. ed. Boston: Beacon. → First published as *The New Belief in the Common Man*. See especially Chapter 5.
- HUITT, RALPH K. 1961 The Outsider in the Senate: An Alternative Role. *American Political Science Review* 55: 566-575.
- KENT, FRANK R. (1923) 1935 *The Great Game of Politics: An Effort to Present the Elementary Human Facts About Politics, Politicians and Political Machines, Candidates and Their Ways, for the Benefit of the Average Citizen*. Garden City, N.Y.: Doubleday.
- KORNBERG, ALLAN 1964 The Rules of the Game in the Canadian House of Commons. *Journal of Politics* 26: 358-380.
- LEITES, NATHAN (1958) 1959 *On the Game of Politics in France*. Stanford Univ. Press. → First published in French.
- LOEWENBERG, GERHARD 1967 *Parliament in the German Political System*. Ithaca, N.Y.: Cornell Univ. Press. → See especially Chapter 4.
- McClosky, HERBERT 1964 Consensus and Ideology in American Politics. *American Political Science Review* 58: 361-382.
- MATTHEWS, DONALD R. 1960 *U.S. Senators and Their World*. Chapel Hill: Univ. of North Carolina Press. → See especially Chapter 5.
- PROTHRO, JAMES W.; and GRIGG, CHARLES M. 1960 Fundamental Principles of Democracy: Bases of Agreement and Disagreement. *Journal of Politics* 22: 276-294.
- TRUMAN, DAVID B. (1951) 1965 *The Governmental Process: Political Interests and Public Opinion*. New York: Knopf.
- VON NEUMANN, JOHN; and MORGENSTERN, OSKAR (1944) 1964 *Theory of Games and Economic Behavior*. 3d ed. New York: Wiley.
- WAHLKE, JOHN et al. 1962 *The Legislative System: Explorations in Legislative Behavior*. New York: Wiley. → See especially pages 141-169, "Rules of the Game," by John C. Wahlke and Leroy C. Ferguson.

#### RUMOR

What are commonly called rumors are those descriptions, forecasts, or explanations of events that are shaped largely in informal encounters and that have not been confirmed by authoritative sources but are, nonetheless, entertained seriously by a substantial portion of an interested collective body. Since the reports are unverified, they may turn out to be either true or false. Although warnings against apocryphal tales can be found in the folklore and literature of many peoples, it is only recently that such tales have become the object of systematic inquiry. Contributions have come from many sources: historians and jurists concerned with the reliability of testimony; psychologists concerned with perception and recall; psychiatrists concerned with the expression of repressed impulses in communicative acts; and sociologists and social psychologists concerned with group problem solving, public opinion, and reactions to disasters. Despite these efforts, comprehension of the phenomenon remains fragmentary, for investigators have faced formidable obstacles in collecting reliable data.

**Two conceptions of rumor.** Pioneering efforts in any field rest on common sense terms, which are vague in reference and contain implicit value judgments, two somewhat different conceptions underlie current studies of rumor. Some regard it as a message that is passed from person to person, assuming that the first speaker in the chain is an eyewitness whose report is accurate and that distortions are introduced in the course of transmission. The basic unit of analysis is the report, and the central problem is to account for its transformation. Since normal communication is assumed to be accurate, rumor is viewed as pathological. Others conceive

of rumor as a summary depiction of an event, a capsule statement that is constructed in the interaction of people who are concerned. For them the unit of analysis is the situation, and the problem is to ascertain how it becomes interpreted. In either case it is useful to distinguish between the *product* and the *process* by which it is shaped.

**Conditions generating rumor.** Investigators are generally agreed that rumors emerge in situations that are inadequately defined—when there has been an unexpected break in the daily routine (the appearance of a stranger), when there is a drastic environmental change (a severe earthquake), when those confronted with alternatives are uncertain of the outcome (an impending election), when people are under sustained tension (a city under siege), or when men performing monotonous tasks become restless but cannot establish the source of their annoyance (rear echelon troops). In each context there is demand for news, although it varies considerably in intensity. When some decision is to be made, any information that might affect the outcome becomes important; once the question has been settled, the sense of urgency wanes, and what had been news becomes history. Ordinarily news is provided by some authoritative source—in modern societies, for example, through the media of mass communication. If demand for news exceeds what is supplied through these channels, interested parties are forced to rely upon one another for information. Thus, any circumstance that hinders the rapid dissemination of news, such as censorship or the physical barriers arising in catastrophes, foment rumors.

Allport and Postman (1947, pp. 33-34) have tried to formalize such observations in their frequently cited formula:  $R \sim i \times a$ ; "the amount of rumor in circulation [R] will vary with the importance of the subject [i] to the individuals concerned times the ambiguity of the evidence [a] pertaining to the topic at issue." Although the exact relationship designated has never been demonstrated, the formula does specify the important variables.

In most communities social contacts are limited by barriers of class, ethnic identity, or other group affiliations, and most rumors are nurtured in well-established informal networks. Festinger and his associates (1948) studied a rumor that a community worker trying to establish a nursery school in a housing project was a communist; they found that 62 per cent of the people with children from one to five years of age had heard it, in comparison to only 28 per cent of the other residents. In an experiment conducted at another housing project, Festinger, Schachter, and Back (1950, pp. 114-

131) planted a rumor that one of the project's special activities was to receive nationwide publicity. They found that this information was conveyed for the most part to individuals who had taken a prominent part in the activity. In disasters, however, information is urgently needed by everyone, and new communication channels emerge spontaneously. One week after the disastrous floods of 1955 a rumor that a dam was about to break sent the populace of Port Jervis, New York, streaming out of the town. In their study of this incident, Danzig, Thayer, and Galanter (see Philadelphia . . . 1958) found that a number of individuals took it upon themselves to broadcast warnings to anyone within range, even though most people turned first to their friends and relatives. During the French Revolution, rumors about brigands employed by aristocrats spread panic through the countryside, and Lefebvre (1932) notes that fugitives, traveling monks, postal carriers, servants sent to warn friends, and local officials all participated in disseminating the alarm. Thus, news is transmitted to persons known to be involved and concerned, and the activation of communication channels varies with the requirements of the situation.

**Content formation.** Although rumors are not necessarily false, neither are they direct descriptions of events. The central problem in the study of rumor is ascertaining how content is formed. This has been approached from three standpoints.

**Experimental studies.** The popular conception of rumor as a message that is passed on from person to person is the basis of experimental studies of serial transmission. The first subject in a series either is asked to describe a picture or is given a predetermined message by the experimenter. He repeats the account to the next subject, who in turn passes on what he has heard. A careful record is made of what happens at each point of contact; additions and omissions are measured. Such studies were first conducted by William Stern in 1902 and have subsequently been repeated by numerous students of legal psychology. The work of Allport and Postman (1947) is the most comprehensive. They summarize their findings in terms of three concepts. Because of *leveling*, the accounts grow shorter, more concise, and more easily grasped. *Sharpening* is the selective perception, retention, and reporting of a limited number of details. *Assimilation* is the tendency of reports to become more coherent and more consistent with the presuppositions and interests of the subjects. Their experiments have been replicated by several others, with substantially the same results; in the course of serial transmission reports become shorter and less



accurate. The conclusion is that rumor content becomes what it is because of faulty perception, limitations of memory, and the inability of most men to repeat, verbatim, complex reports that they have heard.

In spite of the frequent and consistent replications of Allport and Postman's experiments, their conclusions have been challenged by several investigators, who insist that the artificial situation in the experiment differs in important respects from the circumstances under which rumors actually develop. The contention that rumors necessarily become less accurate as they develop has been questioned. In his study of informal communication among American soldiers in the South Pacific during World War II, Caplow (1947) found that many reports became more accurate as they spread. Most men involved in dangerous situations are not gullible; they try to check the accuracy of what they hear, and details that are denied by officials are eliminated. Furthermore, diffusion occurs through a small number of well-defined paths where past rumors had been received with appreciation, and individuals who prove unreliable tend to be eliminated from these networks. Schachter and Burdick (1955) planted a rumor in a girls' school that four students summarily removed from their classroom by the principal were connected with some missing examinations; they likewise found no distortions. On matters in which there is personal involvement, details are not readily forgotten. The contention that rumors necessarily become shorter has also been challenged. Peterson and Gist (1951), who studied the public excitement over the murder of a 15-year-old girl, insist that numerous details accumulate to elaborate a central theme. The prime suspect was believed to be a man who had employed the victim as a baby sitter; and subsidiary reports developed about his absence from the party he was attending at the time of the crime, his relationship with his wife, and other items presumably suppressed by the police. Festinger et al. (1948) also found that various particulars emerged to support the view that the previously mentioned community worker was a communist. Disagreement on this point arises in part from differing conceptions of rumor; individual items may become more concise, but they tend to be supplemented by additional reports. Several studies, then, support Vergil's assertion that *fama crescit eundo*.

*Psychoanalytic approaches.* Psychoanalysts have developed another approach to the problem of rumor content. Since a person taking part in the growth of a rumor can claim he is only passing on

what he has heard, he can speak without assuming full personal responsibility for what he says. This type of activity therefore provides another avenue for releasing repressed impulses. Jung (1910) discusses a rumor, in a girls' school, of sex relations between a teacher and one of the students. The account became disseminated only because of the active participation of a number of students; the girls had similar erotic interests, and the portrayal tapped something that was already "in the air." Distortions are also accounted for in this manner; variations are introduced by participants whose personal needs are slightly different. Rumors take shape, then, as individuals entertain and pass on reports that enable them to give vent to anxieties or hostilities they are reluctant to acknowledge. The method used by psychoanalysts is clinical, but support for their position comes from other studies as well. Allport and Lepkin (1945), for example, found that on the American home front during World War II rumors of waste, special privilege, and corruption in government and in the armed forces were more likely to be accepted by individuals who thought rationing to be unfair or unnecessary, who suffered personal inconveniences from the scarcity of goods, or who had no close friends or relatives in combat zones and were therefore not so personally involved in the war effort.

*Social interaction.* A third approach to the problem of content formation, found largely in the writings of historians and sociologists, is that men who are involved in inadequately defined situations pool their intellectual resources to form reasoned estimates. Rumor is something that is composed in social interaction. Persons deprived of authoritative news speculate about what is happening. They piece together what information they have. Observations are interpreted in the light of what is taken for granted, and the definition that eventually prevails is the one that appears most plausible. For example, rumors developed among German soldiers invading Belgium during World War I that Catholic priests were inciting the populace to commit murder, that machine guns were located in church towers, and that wounded Germans were being mutilated. According to Langenhove (1916), the invaders, who had expected no resistance, were shocked by the guerrilla tactics of the Belgian Army, and reached these conclusions by interpreting events in terms of their conceptions of the French *francs-tireurs* of 1870. Furthermore, as Bysow (1928) points out, rumors tend to become more plausible as they develop. Details that appear incongruent are challenged and eliminated, and

others thought to be relevant are added to make the over-all picture more convincing.

**Patterns of rumor development.** After a comparative analysis of a large number of case studies, Shibutani (1965) concluded that there are several patterns of development. When collective excitement is mild, as in some trivial break in routine, communication occurs through well-established informal networks. Reports are evaluated critically, and the interpretations that are eventually accepted are both plausible and reasonably accurate. On the other hand, when excitement is intense, as in a catastrophe or in the heat of mob action, critical ability is relaxed. New communication channels emerge spontaneously. Rumor content then becomes consistent with the prevailing mood—be it fear, anger, or joy—and men act on the basis of reports that they would question under other circumstances. Rumors are shaped in a selective process, and the basis of selection varies with intensity of collective excitement. Another pattern of development is found in situations characterized by widespread boredom. When demand for accurate information is not urgent, reports are often labeled as rumors and passed on for entertainment by those who do not believe them. These rumors frequently become vehicles for expressing suppressed attitudes.

**Plausibility and consensus.** Before any kind of information can serve as the basis for concerted action, it must be accepted by a substantial portion of the interested public. In many situations people remain skeptical and delay decisions until more authoritative news becomes available; this is especially true of disasters, when many wait until the last possible moment before taking action. Disagreements may arise, and a group may be immobilized. In many cases, however, consensus develops over a rumor. Studies of group problem solving reveal that once a particular view becomes widely accepted, considerable pressure to conform is placed upon those who remain unconvinced. Furthermore, seeing others act on the basis of a rumor tends to make it more credible. Action usually provides an opportunity for reality testing; if there is failure in adjustment, the item is discredited as having been "only a rumor." But many reports are never tested. Beliefs that develop spontaneously in crises are continually being incorporated into the reservoir of popular knowledge to serve as the basis for subsequent judgments.

**Politics and propaganda.** Rumors of all kinds—of secret treaties, troop movements, misdeeds of candidates for public office—have always played an important part in politics; and the problem of

manipulation and control has attracted the attention of scholars. Publicists have sometimes made extravagant claims of successful rumor campaigns, but their contentions are difficult to assess in the absence of independently collected evidence. The propagation of desired rumors is difficult precisely because content does not always remain fixed. News items may be "leaked" into informal networks, but the ultimate product becomes what is plausible to members of the target group. The exploitation of rumors that are already current or the staging of events that lend credibility to them are techniques that have apparently been more successful. During election campaigns and in time of war, there is concern lest opponents sow inconvenient rumors. Denials are usually ineffective, for a rumor does not develop unless it is plausible. Threats of prosecution and other attempts at suppression have also been unsuccessful; indeed, the known existence of censorship tends to create more demand for unofficial news. Appeals for public participation in combating rumors are usually of little avail, although they tend to make people more conscious of the possibility of being deceived. The one procedure that seems to work is the neutralization of sources, for example, the establishment of a conviction that the enemy has a powerful propaganda machine. Then, any inconvenient item can be attributed to the machine and discounted. The techniques of control that actually work are those that rest on principles, in most cases only intuitively recognized, of natural rumor development.

**Explanatory theories.** How are these observations to be explained? For those who regard rumor as pathological, the answer is relatively simple: Accurate communication is precluded by shortcomings of intellect and character of those who make up the links in serial transmission. Most of the other explanatory theories rest on a principle akin to the gestalt law of *Prägnanz*. Prasad (1935, p. 7) writes that as long as a change in the environment remains unaccounted for, there is a distracting sense of incompleteness; this elicits an attempt to understand the new circumstances by completing the incomplete. Allport and Postman (1947, pp. 36–38) declare that a report enjoys dissemination if it permits a "good closure"—if it allows relief from tension, justifies emotions that are unacceptable, or makes the world more intelligible. Festinger (1957) has attempted a more formal statement in his theory of cognitive dissonance; the existence of inconsistency is uncomfortable, and a person will (a) try to reduce it and (b) actively avoid situations and information that are likely to increase it. Rumors, then, are neither the



product of faulty memory nor of efforts to defraud; most of them represent honest attempts to comprehend ambiguous situations.

Although many have been attracted to the study of rumor by concern over accurate communication, the significance of the subject matter extends far beyond this. It adds to an understanding of the local diffusion of materials presented on the media of mass communication. It also reveals the limitations of formal social control. Whenever men distrust official sources, they rely upon one another for news; all authoritarian regimes are plagued by rumors. Furthermore, some of the ways in which popular beliefs develop are disclosed. Those who believe that rational conduct always rests on verified knowledge can acquire a more realistic view of the premises underlying our daily rounds. Since rumor is a form of group problem solving, it is also of importance to students of social change. Far from being pathological, rumor is an integral part of the social process, an important aspect of the continuing efforts of men to cope with the vicissitudes of life.

TAMOTSU SHIBUTANI

[Other relevant material may be found in COGNITIVE THEORY; COMMUNICATION; GESTALT THEORY; PROPAGANDA.]

#### BIBLIOGRAPHY

- ALLPORT, FLOYD H.; and LEFKIN, MILTON 1945 War-time Rumors of Waste and Special Privilege: Why Some People Believe Them. *Journal of Abnormal and Social Psychology* 40: 3-36.
- ALLPORT, GORDON W.; and POSTMAN, LEO 1947 *The Psychology of Rumor*. New York: Holt.
- BYSOW, L. A. 1928 Gerüchte. *Kölner Vierteljahrshefte für Soziologie* 7: 301-308; 416-426.
- CAPLOW, THEODORE 1947 Rumors in War. *Social Forces* 25: 298-302.
- FESTINGER, LEON 1957 *A Theory of Cognitive Dissonance*. Evanston, Ill.: Row, Peterson.
- FESTINGER, LEON et al. 1948 A Study of a Rumor: Its Origin and Spread. *Human Relations* 1: 464-485.
- FESTINGER, LEON; SCHACHTER, STANLEY; and BACK, KURT (1950) 1963 *Social Pressures in Informal Groups: A Study of Human Factors in Housing*. Stanford (Calif.) Univ. Press.
- JUNG, CARL G. (1910) 1916 A Contribution to the Psychology of Rumor. Pages 176-190 in Carl G. Jung, *Analytical Psychology*. Authorized translation edited by Constance E. Long. New York: Dodd.
- LANGENHOVE, FERNAND VAN 1916 *The Growth of a Legend: A Study Based Upon the German Accounts of Francs-tireurs and "Atrocities" in Belgium*. New York: Putnam. → Simultaneously published as *Comment naît un cycle de légendes*.
- LEFEBVRE, GEORGES 1932 *La grande peur de 1789*. Paris: Colin.

PETERSON, WARREN; and GIST, NOEL P. 1951 Rumor and Public Opinion. *American Journal of Sociology* 57: 159-167.

PHILADELPHIA, INSTITUTE FOR RESEARCH IN HUMAN RELATIONS 1958 *The Effects of Threatening Rumor on a Disaster-stricken Community*. By Elliott R. Danzig, Paul W. Thayer, and Lila R. Galanter. Disaster Research Group, Disaster Study No. 10. Washington: National Academy of Sciences-National Research Council.

PRASAD, JAMUNA 1935 The Psychology of Rumour: A Study Relating to the Great Indian Earthquake of 1934. *British Journal of Psychology* 26: 1-15.

SCHACHTER, STANLEY; and BURDICK, HARVEY 1955 A Field Experiment on Rumor Transmission and Distortion. *Journal of Abnormal and Social Psychology* 50: 363-371.

SHIBUTANI, TAMOTSU 1965 *Improvised News: A Sociological Study of Rumor*. Unpublished manuscript.

## RUNS

See under NONPARAMETRIC STATISTICS.

## RURAL SOCIETY

The scientific study of rural society as a specialized area of sociology is a development of the twentieth century and prior to World War II had its growth principally in the United States. Since 1950 such study has developed institutional support in many countries. The central concern in the sociological study of rural society is with its social organization—the social systems (or subsystems) and their interrelationships within rural society, with urban society, and within the total society. This study has been approached with ecological, cultural, and behavioral emphases. Demographic analysis has had an important place. Some areas of rural sociological analysis are closely related to the concerns of other social science disciplines; for example, settlement patterns with human geography and land tenure with agricultural economics. The adoption and diffusion of technological innovations is one of the areas to which rural sociologists have given special attention.

Sociological investigation of rural society may be classified into four broad categories: studies using rurality as the independent variable; comparative studies of rural societies; studies for which rural society is the setting within which family systems, ecological systems, cultural systems, personality systems, and other phenomena are analyzed; and studies of social change within each of the previous three categories.

A major attempt to generalize on rural-urban differences that are repeated in time and space was made by Sorokin and Zimmerman in 1929, drawing upon the comprehensive compilation of then

existing knowledge and theories and research in European, Asiatic, and American scientific literature represented in *A Systematic Source Book in Rural Sociology* (Sorokin et al. 1930-1932). They hypothesized that the differentiation between rural and urban does not tend to increase perpetually but, rather, that the historical curve of the differentiation shows something similar to a parabolic curve without complete obliteration of all differences (Sorokin & Zimmerman 1929, pp. 608-636).

No comparable comprehensive compendium of the much more numerous and more sophisticated studies completed since the 1920s is currently available. Summarizations for major substantive areas of sociological research are represented for North America by the papers given at the 1961 annual meeting of the Rural Sociological Society (Copp 1964); for the major regions of the world by the papers prepared for the First World Congress for Rural Sociology (1964); and by monographic studies, such as those on the adoption and diffusion of ideas in agriculture (Lionberger 1960; Rogers 1962).

**Importance of rural society.** Despite the ever-accelerating rate of urban growth throughout the world, especially of large cities, since 1800, a majority of the world's population is rural and a majority of the nations of the world have predominantly rural populations. The percentage of the world's population living in localities of less than 20,000 was 97.6 in 1800, 95.7 in 1850, 90.8 in 1900, and 79.1 in 1950 (United Nations, Bureau of Social Affairs, 1957, pp. 111-143). For major world areas, the estimates of the percentage of the population living in places of less than 20,000 in 1950 were: Africa, 91; Asia, 87; Central America, 79; South America, 74; U.S.S.R., 69; Europe (excluding the U.S.S.R.), 65; North America, 58; and Oceania, 53 (Joint UN/UNESCO Seminar on Urbanization in the ECAFE Region, Bangkok, 1956, 1957, p. 98). This percentage decline in rural population has been accompanied by an increase in the number of persons living in localities under 20,000; the estimated number was 1,897.8 million in 1950, compared with 1,460.1 million in 1900. In the majority of the less-developed countries, the rural population has continued to grow in numbers, whereas in the higher-income countries the pattern has varied in recent decades.

The following illustrations use individual countries' definitions of rural and urban. In England and Wales, which became less than half rural by 1851, the equilibrium at about 20 per cent rural maintained since 1921 is associated with some increase in the number of rural residents because the

total population has grown (Saville 1957). In the United States, which became less than 50 per cent rural by 1920 and was 30 per cent rural in 1960, the rural population continued to grow in numbers until 1940 and since then has remained relatively constant, despite the reducing effects of changed statistical procedures. In Brazil, which contains one-third of all the people of the 20 Latin American countries, the percentage rural decreased from 68.8 in 1940 to 54.9 in 1960; at the same time, the number of rural people increased by more than 10 million (Miró 1964). In India, although the urban population has grown at an accelerating rate, 82 per cent of the total population was rural in 1961; the rural population increased by 61 million during the preceding decade, whereas the urban population gained 16 million (Taylor et al. 1965, pp. 104-105). The concentration of rural population in low-income countries and the importance of rural society in plans for economic and social development give added significance to understanding rural society and its interrelationships with urban society.

**The functions of rural society.** The production of food and other raw materials is a basic function of rural societies; indeed, in modern society the survival of the urban sector is dependent upon the effective conduct of this function. Historically, rural society has also functioned as a supplier of people to the city; the relative importance of this rural-to-urban migration is probably greatest in societies that are undergoing modernization. A third distinctive function of rural society is its custodianship of natural resources. Fourth, modern rural societies increasingly serve a residential function for persons with urban-centered sources of livelihood. Finally, in times of societal crisis, there has been a tendency for rural society to perform a "security" function, as represented by urban-to-rural or postponed rural-to-urban migration. In addition, economic functions ancillary to or independent of the primary or "field" activities are characteristically found in rural areas in some form and in some degree but are not considered as distinctive functions of rural society. Rather, the type and extent of these nonprimary functions vary widely among countries and over time.

### The meaning of rural

A uniform definition of the concept "rural" is crucial if sociological analysis is to result in valid generalizations. In the past decade, increased attention has been given to considering the minimum and essential criteria of rural society. Recent research has found that some characteristics once



advanced as attributes of rural society may also be found in the city; close personal primary-group relations are one example. The interchange of people between rural and urban areas through large-scale migration, the influence of urban-centered mass media, greater interdependence of rural and urban economies, and other types of increased systemic linkage are believed to have reduced rural-urban differences. Also, research revealing the diversity within both rural and urban societies has led to a re-examination of the rural concept.

Although there is a broad general consensus that the term "rural" refers empirically to populations living in areas of low density and to small settlements, there are wide variations in the cutting point used operationally to distinguish rural from urban. There is even less agreement on the qualitative attributes that distinguish a rural from an urban community or society. "Rural" has been defined in terms of one or more empirical attributes, as a constructed type with qualitative attributes, and by the two in combination. The methodological and theoretical problems in the definition of the concept "rural" are similar to those in the definition of "urban" and in the use of constructed types in general. Occupational, demographic, ecological, social organizational, and cultural characteristics have all been used as defining attributes.

In the highly influential formulation by Sorokin and Zimmerman (1929), the principal criterion was occupational—the collection and cultivation of plants and animals. They argued that from the agricultural occupation there followed a series of other relatively constant qualitative and quantitative differences between rural and urban communities, most of which were causally connected or interrelated. The use of occupation as the defining criterion for rural society has not found general acceptance by sociologists. Widespread social changes have frequently resulted in farmers living intermingled with residents of the countryside who follow nonfarm occupations at urban places of work, even to the extent that farmers are a minority within their local rural community. Also, the farm operator or members of his household may have a nonfarm occupation. Sociologically, the use of occupation as the criterion leads to a greater interest in the aggregate of individuals than in the properties of a social system.

The most general current practice is to use two demographic variables—absolute size and density of settlement—in defining "rural." In most countries the dividing line between rural and urban is set at population aggregates of somewhere between 1,000 and 5,000 inhabitants. There is a high level

of agreement that the conditions of social action and social organization set by low population density and small size of community have the following effects. First, the pattern of social interaction involves relatively few contacts per individual in a given unit of time and provides relatively little opportunity for personal anonymity; second, there is relatively little division of labor, as there is a restricted range of occupations and a limited variety of social roles; and thus there is relatively little heterogeneity at the community level as compared with urban society. Social roles tend to be more diffuse and less segmentalized, along with being less impersonal. Moreover, social class identification at the community level has a tendency to be based more on ascribed and personal attributes than on general and achieved attributes; from this fact and that of the limited variety of occupational and social roles, it follows that the limits of social mobility are more restricted within rural society. Finally, because of less anonymity, normative deviancy within the rural community tends to be less frequent than in the city.

The evidence that there are universal differences in cultural characteristics associated with differences in density and size is not convincing. An alternative proposal is that "rural" be defined in terms of a particular pattern of value configurations or some other cultural attributes (Bealer et al. 1965). However, since such a construct would make rural society independent of a specific territorial space, it has had little application. Still other suggested criteria for characterizing rural society include: relative size and density rather than arbitrary limits; the ecological notion of relative isolation from or accessibility to larger population centers, expressed by Stewart (1958) in terms of a social network map deriving the proportion of urban contacts from objective indices; and another ecological alternative whereby people engaged in "field" activities would be classed as rural and those engaged in "center" activities would be termed urban. "Field" activities are considered those involving the production of foods, fibers, ores, and raw materials and requiring high man-land ratios (Loomis & Beagle [1950] 1955, p. 204).

The statement by Hatt and Reiss ([1951] 1957, pp. 19–21) that "there are considerable differences among the criteria and the ease with which they can be used in research, the degree to which they provide meaningful theoretical and empirical differentiation, their constancy in time and space, their cultural relativity, and their causal connection or interrelationship" applies to both rural and urban communities. The size and density criteria can be

applied if the focus of analysis of rural society is on demographic variables or on social organization. Dissatisfaction with these two criteria seems to be associated particularly with attempts to generalize about cultural systems, including values, and about personality systems.

The idea of a rural-urban continuum is now generally preferred to a rural-urban dichotomy. In research operations, however, particularly for population analysis, the use of a dichotomy is a common practice. Or, as in the procedure followed by the United States Bureau of the Census, a trichotomy may be set up by dividing the rural into two parts—farm and nonfarm. That the rural-urban continuum concept may need to be reinterpreted is suggested by the Duncan and Reiss finding (1956, p. 40) that, for 11 size-of-community groups, equal differences in the numerical size of two communities were in no case associated with equal differences in the index of any of the population characteristics examined. It may be that unequal rather than equal size-of-community intervals may more closely approximate the threshold of significant differences [see COMMUNITY-SOCIETY CONTINUA].

Within rural society there are wide variations in social organization and in cultural systems. Such types as "primitive," "nomadic," and "peasant" are illustrations. Feudal social organization has been identified as rural and also has been advanced as an ideal type opposed to both rural and urban (Sjoberg 1952). A number of constructed typologies of societies have been formulated, such as "Gemeinschaft-Gesellschaft" and "sacred-secular," which have mistakenly been used as synonymous with rural-urban. Such constructed typologies may, however, be appropriately understood as intersecting the rural-urban typology.

The *traditional-modern* construct, as advanced by Lerner (1958) and adapted by others (Larson & Rogers 1964), seems to have particular promise for the analysis of contemporary rural society. The crucial point here is that members of modern social systems view new ideas more favorably than do members of traditional systems; that is, they differ in their empathic capacity. In addition, modern social systems are characterized by a relatively highly developed technology with a complex division of labor, a high level of literacy and formal education, cosmopolitan social relationships, a relatively large proportion of secondary social relationships, and an emphasis upon economic rationality. Further, the modern society has a high level of development of communication media, which give access to the cumulative pool of knowledge of the larger society and provide a comparatively high

degree of linkage and articulation among the subsystems in relation to the societal system.

### Social change in rural society

The trend toward modernization of traditional rural societies, which has been under way for well over a century in North America, Great Britain, and western Europe, has accelerated and become more pervasive during the twentieth century, especially since World War II. Rural societies of the modern type are so interlinked with their national systems and the international system that the changes in their social organization are highly similar in many respects. Traditional-type rural societies are generally moving toward the same type of linkage. The process by which a traditional, largely isolated and segmented, rural society becomes a modern-type rural society integrated with the national society is a special case of the more general process of societal change.

Although there are complex theoretical and methodological problems involved in the analysis of social change, there is considerable agreement about the direction and universality of major changes that have been under way in social and cultural systems. These major changes have been expressed in such highly general terms as increased structural differentiation, or specialization, of social systems; more universalistic norms; and the growth of a rational orientation (Moore 1963; Parsons 1964). This discussion of social change will focus on the following selected factors, which have special significance for the rural social changes of the present century: the growth and application of technological and scientific knowledge; the increase in urbanization; the growth of the money-and-market economy; and the increase in purposive innovation, reflecting rationality and a favorable attitude toward change. These factors may operate partly independently and partly in association and may be in part and in some ways circular in their effects. In large measure, they operate generally as external forces for change in rural societies and their social systems. The changes which these forces generate within rural communities in turn produce additional internal community changes.

Empirical data for use in describing the current situation, measuring changes, and making comparative studies are most adequate for the more modern rural societies, such as those in North America, western Europe, and Japan, but even for these there are major deficiencies. In general, the available information on the rural societies of the world is less complete regarding social organization and the different categories of social systems than



it is for population, technology, physical resources, and production, although these data may be reflective of social organization variables.

**Technological and scientific advances.** Agricultural production technology offers some of the most dramatic examples of the social consequences of technological change through major innovations and improvements (Moore 1963, p. 77). The extent to which technological advances and science-based practices have been diffused and adopted varies widely from country to country. Even in the most modernized nations, survival of traditional methods may be found in some local areas. It is estimated that about 90 per cent of the power used on the world's farms is still supplied by human beings and animals.

In 1953 the Woytinskys estimated that more than half of the world's labor force was engaged in agriculture (Woytinsky & Woytinsky 1953, p. 364). The proportion so engaged ranged from about 5 per cent in a modernized food-importing nation, such as Great Britain, to more than 80 per cent in more traditional societies. The general trend is for the proportion of the labor force engaged in agriculture to decline; for example, in the 20 Latin American countries the percentage fell from 61 in 1936 to 47 in 1960 (United Nations, Economic Commission for Latin America 1964, p. 23). The percentage decline may be accompanied by an increase in the number of persons in agriculture, as in Latin America, or by a decrease, as in the United States.

The effects of the application of technological and scientific knowledge in the rural sector of a modern society can be seen in the United States. Trend data for 1870-1960 show that productivity per worker in agriculture and per man-hour increased rather regularly at slightly less than 1 per cent per year until the mid-1930s, after which it increased at the accelerated rate of 4 to 5 per cent annually (Meiburg & Brandt 1962). Between 1940 and 1959, while total farm production increased by 52 per cent, the proportion of land allocated for farming and the number of harvested acres changed only slightly and the number of farm workers and aggregate man-hours of farm work declined to less than half of the peak reached prior to 1920 (Larson & Rogers 1964). The farm population correspondingly declined to less than half of its peak number and became a minority within the rural population. Increased worker productivity was accompanied by fewer but larger farms. The number of farms per 1,000 acres of land decreased from 7.5 in 1880 to 3.3 by 1959. Such a reduction led to a decrease in population density in all areas

where farm population losses were not offset by growth of the rural nonfarm population. At the local level, the territoriality and size conditions for many rural social systems were changed as a consequence of the series of changes in farming.

Modern transportation has served to increase the frequency of rural-urban interchange and to facilitate the use by rural people of a network of centers with specialized functions. Technological developments in communication also have a major influence in the modernization process by greatly enlarging and facilitating access to the pool of knowledge. They support the linkage of the rural community and its subsystems into a network of systems that may cross national boundaries, support centralization of decision making, foster a broader "world view" on the part of the rural resident, and change the nature of the interaction process in communication from oral, direct, and two-way to mediated or one-way. For example, the adoption of radio has been rapid since 1950, even in the isolated areas of the more traditional societies (United Nations, Bureau of Social Affairs, 1957, p. 79). The mass media are largely urban centered and urban controlled, extending their influence without requiring the rural resident to make physical contact with the city. To date, studies of the diffusion of new ideas that may be adopted by individuals voluntarily, such as farm practices, indicate that the mass media are important in creating awareness and interest but that personal contacts are a more important information source at the adoption stage (Bohlen 1964, pp. 281-283). Furthermore, different types of adopters make differential use of mass media; the innovators make the most use of it, whereas the slow adopters place more reliance on personal and primary-group information sources.

**Urbanization.** In recent decades the rate of urbanization has been declining in the more modernized societies, where urbanization is now greatest, and generally increasing in more traditional rural societies. In the latter, the rapid urban expansion not only reflects the normal "pull" factors but also is the consequence of powerful "push" factors in rural areas with heavy population pressure on land resources and limited nonagricultural development (United Nations, Bureau of Social Affairs, 1957, pp. 124-125). The extent to which a high level of urbanization may provide an outlet for excess rural population is suggested by the estimate of Beale et al. (U.S. Department of Agriculture, Economic Research Service, Economic and Statistical Analysis Division, 1964a) that as of 1958 in the United States about 63 per cent of the farm-born adults were nonfarm residents.

Net rural-to-urban migration is selective as to who leaves and who remains. Migration to cities tends to select young adults, although there may be instances where such selectivity does not hold (Davis & Golden 1954-1955). Generally, in the more modernized societies relatively more females than males migrate from rural areas; in the more traditional societies, more males leave. However, there are exceptions to this sex selectivity pattern. In a country as modernized as Japan, relatively more rural males than females go to the city; whereas in the Latin American countries, rated as more traditional on the whole, the opposite holds true.

**Money-and-market economy.** The Woytinskys estimate that in the world as a whole not more than one farmer in five is engaged in commercial farming, producing largely for the open market; the balance are subsistence farmers, producing mainly to meet their own needs (Woytinsky & Woytinsky 1953, p. 452). In the countries with more traditional societies where Western colonization occurred, the development of large-scale commercial agricultural production and exploitation of mineral and oil resources for the world market has produced a form of dual economy [see ECONOMY, DUAL].

In the modernized societies the money-and-market economy puts a premium upon rationality and efficiency, which has led to increased specialization of agricultural enterprises by areas and by individual farmers, an increased reliance on purchased inputs in farm production, the development of new types of credit systems for farmers, the growth of farmers' marketing and purchasing co-operatives, and major changes in the marketing system for farm products.

**Planned change.** During the present century, and especially during the past two or three decades, national governments have had a conspicuous role in beginning innovations. The United Nations complex of international organizations and such regional arrangements as the Organization of American States, as well as the United States—and more recently a few other national governments—have developed extensive programs aimed at helping the underdeveloped nations to speed up the modernization of their traditional rural societies. Practically all nations with quite modernized rural societies have undertaken a wide range of activities to equalize social conditions. Illustrations are price and income support programs for farmers, the development of infrastructure (improved roads, electric power lines, telephones, and water supply systems), rural mail delivery, improvement of rural

educational facilities, and special programs for library service, medical care, and hospitals. Action for the purpose of maintaining and strengthening the national system ranges from relatively limited land reform, resettlement, and land consolidation in such countries as Italy, the Netherlands, and Sweden to attempts at inclusive restructuring of rural society in the U.S.S.R. and mainland China.

In at least the more modern societies, the net effect of these and related factors is to bring several of the social systems of which rural people are members under heavy pressure to restructure, both horizontally and vertically. The rise in standards and levels of expectation shared by rural people affects both "public" and "private" systems. Realignments are required in order to provide the necessary population and financial base for supplying the more specialized services at reasonable cost. Thus, the territoriality of systems providing education, health care, local government, and numerous economic functions for rural people is under strain to readjust to the patterns made possible by new transportation technology. The reduction of the number of public school districts in the United States from 127,000 in 1932 to 49,000 in 1958 through consolidation is illustrative (R. M. Williams 1964, p. 23). There is an over-all trend toward mutual interdependence among locality groups, with different functions located at different centers, with rural families using a number of centers for services, and with each type of economic and social service tending to have its own unique service area. An increase in the heterogeneity of the rural community results from the decrease in the percentage of the population engaged in farming or in other "field" activities, the separation of place of work from residence for the commuting nonfarm workers whose employment is outside the community, and—in some instances—the increase in seasonal residence by urban dwellers.

Urban-centered and rural-centered systems are increasingly in contact, either cooperating formally or informally with greater systemic linkage or in conflict to maintain system boundaries. Urban expansion puts especial pressure on rural areas at the growing edge of the city.

Many of these social changes occur unevenly, even within countries. There tend to be regional differences, community differences, and variations by class status and by culturally deviant groups. Rural communities with expanding populations have been shown to have adjustment problems different from those with stable or declining populations (Larson & Lutz 1961).



Ecological spatial relationships possess a considerable degree of built-in structural stability; land use, field, and settlement patterns are not easily changed (W. M. Williams 1964). Outside the areas of wholesale land reform or new colonization, the spatial arrangements in rural areas have changed more slowly than technology or some of the social systems. However, there have been several notable ecological changes: land consolidation programs and limited transfer of farmers from villages to dispersed farms in western Europe; enlarged farms generally and some shift of farmers to villages in the Great Plains wheat belt of the United States; and the thickening of settlement within the commuting area of urban centers.

### Contemporary rural-urban comparisons

The pattern of higher fertility in rural than in urban areas is observed in practically all countries for which data are available, although it seems generally to be least marked within the most traditional societies. The differentials have been persistent, even in the more modernized nations. In 1960, the number of children ever born per 1,000 rural women aged 35-39 years in the United States was more than one-fifth larger than for urban women; in none of the 50 states was the urban fertility as high or the average size of family as large as the rural (U.S. Department of Agriculture, Economic Research Service, Economic and Statistical Analysis Division, 1964b). Because of the heavy out-migration of the rural born, the annual rate of growth is higher for the urban than for the rural population in nearly all nations for which data are available for the most recent decades.

The combined effects of differential fertility and differential migration leave the rural areas with a disproportionate burden of dependents, especially children, and a relative shortage of persons in the productive adult years. With the exception of certain Asian and African countries, rural areas have more males than females—particularly at the younger adult ages—because of the selective factor in net migration. These differences in sex ratios reflect the difference in work roles. In the most modernized countries, the greatest opportunity for gainful employment for females is in nonfarm work; the "field" activities are predominantly men's work.

In a modern society, the rural-farm sector is the only one in which place of work, place of residence, occupation, and family are so tightly wrapped together. Sorokin and Zimmerman's generalization (1929, p. 334) that the rural family is more stable than the urban family has been supported by some

recent studies in the United States (see Burchinal 1964, pp. 177-178). However, in some instances, as for Panama and Venezuela in 1950, a significantly higher proportion of consensual unions exist in rural areas (Miró 1964). That role performance by sex and age differs between rural and urban families in some respects is suggested by research on the United States; however, according to Burchinal (1964), there is evidence that rural and urban family organizations in the United States are becoming more alike.

There are major gaps in comparative rural-urban cross-national research on most of the social subsystems and particularly in group-level research that uses the concrete social system—such as the school or the church—as the unit of analysis. However, to some extent, inferences about these systems may be derived from the more available demographic, occupational, and similar census-type and aggregate data. The high association of illiteracy with rurality in national comparisons and the fact that the pattern of fewer years of school completed by rural adults has persisted in a nation such as the United States suggest rural-urban differences in the systems of formal education. The evidence shows quite consistently that, at a minimum, educational and other social-service type systems in rural areas tend to be smaller in size and less specialized in function. The extent of the structural influence of the conditions for social action in rural society is hinted at by research showing that youth clique groups are smaller in rural than in urban areas (Hare 1962, p. 225).

The occupational influence of agriculture contributes to the persistence of distinctive characteristics in rural society. The alliance of occupation, residence, and family has already been noted. In addition, although there are areas where large estates, concentrations of ownership, or large-scale state or collective units may be found, farming generally is characterized by small units of operation, with the management and worker roles combined in the farmer and with the work performed primarily by the farmer and members of his family; so the ratio of independent to other workers is in sharp contrast to that in the nonagricultural pursuits. Occupational self-recruitment is greater for farming than for the other major occupational categories; that is, a higher percentage had fathers in the same occupation. Finally, studies covering both Western and non-Western countries, some over several decades, show that income of farmers is consistently and persistently below that of nonfarmers (see Bellerby 1956).

A number of studies have shown that demo-

graphic, occupational, and economic factors have a gradient pattern in relation to large urban centers, varying quite consistently with distance. Thus, the youth dependency ratio and the fertility ratio in the white rural-farm population rose regularly and consistently with increasing distance from the nearest standard metropolitan statistical area in the United States in 1960 (Beegle 1964).

Rural-urban differentials in modernized societies have been greatly reduced or even eliminated in many respects. Although an increase in cultural homogeneity is suggested, differences in social organization persist. The several theories about community growth, center functions, and social differentiation generally support the proposition that the conditions of social action result in the persistence of rural-urban differences in scale and complexity over a wide range of social systems. Purposive action is directed at minimizing the differences in those systems vested with public concern.

OLAF F. LARSON

[See also AGRICULTURE; MODERNIZATION; POPULATION, article on POPULATION DISTRIBUTION.]

#### BIBLIOGRAPHY

- BEALER, ROBERT C.; WILLITS, FERN K.; and KUVLESKY, WILLIAM P. 1965 The Meaning of "Rurality" in American Society: Some Implications of Alternative Definitions. *Rural Sociology* 30:255-266.
- BEEGLE, J. A. 1964 Population Changes and Their Relationship to Changes in Social Structure [in the Rural Areas of America]. *Sociologia ruralis* 4:238-252.
- BELLERBY, JOHN R. 1956 *Agriculture and Industry Relative Income*. New York: St. Martins.
- BOHLEN, JOE M. 1964 The Adoption and Diffusion of Ideas in Agriculture. Pages 265-287 in James H. Copp (editor), *Our Changing Rural Society: Perspectives and Trends*. Ames: Iowa State Univ. Press.
- BRACEY, HOWARD E. 1959 *English Rural Life: Village Activities, Organisations and Institutions*. London: Routledge.
- BURCHINAL, LEE G. 1964 The Rural Family of the Future. Pages 159-197 in James H. Copp (editor), *Our Changing Rural Society: Perspectives and Trends*. Ames: Iowa State Univ. Press.
- COPP, JAMES H. (editor) 1964 Projection Papers: Orientations for Rural Sociological Research and Action. Unpublished manuscript, University Park, Pa., Rural Sociological Society. → Papers given at the 1961 annual meeting of the Rural Sociological Society.
- DAVIS, KINGSLEY; and GOLDEN, HILDA H. 1954-1955 Urbanization and the Development of Pre-industrial Areas. *Economic Development and Cultural Change* 3:6-26.
- DUNCAN, OTIS DUDLEY; and REISS, ALBERT J. JR. 1956 *Social Characteristics of Urban and Rural Communities, 1950*. U.S. Bureau of the Census, Census Monograph Series. New York: Wiley.
- HARE, ALEXANDER P. 1962 *Handbook of Small Group Research*. New York: Free Press.
- HATT, PAUL K.; and REISS, ALBERT J. JR. (editors) (1951) 1957 *Cities and Society: The Revised Reader in Urban Sociology*. 2d ed., rev. Glencoe, Ill.: Free Press.
- JOINT UN/UNESCO SEMINAR ON URBANIZATION IN THE ECAFE REGION, BANGKOK, 1956 1957 *Urbanization in Asia and the Far East*. Edited by Philip M. Hauser. Calcutta: UNESCO Research Centre on the Social Implications of Industrialization in Southern Asia.
- LARSON, OLAF F.; and LUTZ, E. A. 1961 Adjustments in Community Facilities Taking Place and Needed. Pages 285-336 in National Committee on Agricultural Policy, *Adjustments in Agriculture: A National Basebook*. Edited by Mervin G. Smith and Carlton F. Christian. Ames: Iowa State Univ. Press.
- LARSON, OLAF F.; and ROGERS, EVERETT M. 1964 *Rural Society in Transition: The American Setting*. Pages 39-67 in James H. Copp (editor), *Our Changing Rural Society: Perspectives and Trends*. Ames: Iowa State Univ. Press.
- LERNER, DANIEL 1958 *The Passing of Traditional Society: Modernizing the Middle East*. Glencoe, Ill.: Free Press. → A paperback edition was published in 1964.
- LIONBERGER, HERBERT F. 1960 *Adoption of New Ideas and Practices*. Ames: Iowa State Univ. Press.
- LOOMIS, CHARLES P.; and BEEGLE, J. ALLAN (1950) 1955 *Rural Social Systems: A Textbook in Rural Sociology and Anthropology*. London: Bailey & Swinfen.
- MEIBURG, CHARLES O.; and BRANDT, KARL 1962 *Agricultural Productivity in the United States: 1870-1960*. Stanford University, Food Research Institute, *Food Research Institute Studies* 3:63-85.
- MIRÓ, CARMEN A. 1964 The Population of Latin America. *Demography* 1, no. 1:15-41.
- MOORE, WILBERT E. 1963 *Social Change*. Englewood Cliffs, N.J.: Prentice-Hall.
- PARSONS, TALCOTT 1964 Evolutionary Universals in Society. *American Sociological Review* 29:339-357.
- ROGERS, EVERETT M. 1962 *Diffusion of Innovations*. New York: Free Press.
- SAVILLE, JOHN 1957 *Rural Depopulation in England and Wales: 1851-1951*. London: Routledge.
- SJOBORG, GIDEON 1952 Folk and "Feudal" Societies. *American Journal of Sociology* 58:231-239.
- SOROKIN, PITIRIM; and ZIMMERMAN, CARLE C. 1929 *Principles of Rural-Urban Sociology*. New York: Holt.
- SOROKIN, PITIRIM A.; ZIMMERMAN, CARLE C.; and GALPIN, CHARLES J. (editors) 1930-1932 *A Systematic Source Book in Rural Sociology*. 3 vols. Minneapolis: Univ. of Minnesota Press.
- STEWART, CHARLES T. JR. 1958 The Urban Rural Dichotomy: Concepts and Uses. *American Journal of Sociology* 64:152-158.
- TAYLOR, CARL C. et al. 1965 *India's Roots of Democracy*. Calcutta: Orient Longmans.
- UNITED NATIONS, BUREAU OF SOCIAL AFFAIRS 1957 *Report on the World Social Situation*. New York: United Nations.
- UNITED NATIONS, ECONOMIC COMMISSION FOR LATIN AMERICA 1964 *The Economic Development of Latin America in the Post-war Period*. New York: United Nations.
- U.S. DEPARTMENT OF AGRICULTURE, ECONOMIC RESEARCH SERVICE, ECONOMIC AND STATISTICAL ANALYSIS DIVISION 1964a *Characteristics of the U.S. Population by Farm and Nonfarm Origin*. By Calvin L. Beale, S. C. Hudson, and Vera J. Banks. Agricultural Economic Report No. 66. Washington: The Department.



- U.S. DEPARTMENT OF AGRICULTURE, ECONOMIC RESEARCH SERVICE, ECONOMIC AND STATISTICAL ANALYSIS DIVISION 1964b *A Summary of Selected Characteristics of the Urban and Rural Populations, by State, 1960*. Washington: The Department.
- WILLIAMS, ROBIN M. JR. 1964 American Society in Transition: Trends and Emerging Developments in Social and Cultural Systems. Pages 3-38 in James H. Copp (editor), *Our Changing Rural Society: Perspectives and Trends*. Ames: Iowa State Univ. Press.
- WILLIAMS, W. M. 1964 Changing Functions of the Community. *Sociologia ruralis* 4:299-314.
- WORLD CONGRESS FOR RURAL SOCIOLOGY, FIRST 1964 [Proceedings.] *Sociologia ruralis* 4, no. 3/4.
- WOYTINSKY, WLADIMIR S.; and WOYTINSKY, EMMA S. 1953 *World Population and Production*. New York: Twentieth Century Fund.

## RUSH, BENJAMIN

Benjamin Rush (1746-1813), American physician, politician, and social reformer, was born in Byberry, Pennsylvania, the son of a farmer and gunsmith. After graduating from the College of New Jersey (later Princeton) in 1760, he was apprenticed to the physician John Redman. He entered the University of Edinburgh in August 1766 and received his medical degree in June 1768, after presenting a thesis, *De coctione ciborum in ventriculo*, based on several experiments performed on himself and a friend. Shortly after his return to Philadelphia in 1769, Rush was elected professor of chemistry in the College of Philadelphia (now the University of Pennsylvania) and began the practice of medicine.

Like his contemporaries Franklin and Jefferson, Rush was a representative of the American Enlightenment. As a republican, he was an ardent proponent of American independence. In 1776 he was made a member of the Continental Congress and signed the Declaration of Independence. He joined the Continental Army in December 1776 and from April 1777 to January 1778 served as surgeon-general of the armies of the Middle Department. But differences arose between Rush and William Shippen, the director-general, which led to Rush's resignation and resumption of private practice in Philadelphia. In 1789 he was appointed professor of the theory and practice of medicine in the medical faculty of the College of Philadelphia.

At the same time Rush continued his social and political activities. He felt obliged to use his mind and pen in any way that would advance national unity and foster political stability in the young American republic. Thus, he advocated the national constitution of 1787 and helped to frame the Pennsylvania state constitution of 1790. From

1797 to 1813 he served as treasurer of the United States Mint. Moreover, recognizing the relationship between an informed public opinion and popular sovereignty, Rush saw the need to prepare the citizens for the task of governing. Consequently, he advocated a national system of education, with various state-supported schools, including a national university and technical schools. In addition, Rush also championed higher education for women, and he was largely responsible for the founding of Dickinson College in western Pennsylvania.

But Rush's efforts for the public good went beyond education. His major impulsion to action arose from the desire to improve the conditions of men in a rational manner. He was an early and passionate proponent of the abolitionist cause, serving as president of the Pennsylvania Society for Promoting the Abolition of Slavery. He advocated temperance and legal control of drinking. Penal reform was another cause that Rush made his own, and in 1787, together with Franklin and several like-minded men, he organized the Philadelphia Society for Alleviating the Miseries of Public Prisons. It was this group that succeeded in having the death penalty in Pennsylvania abolished for all crimes except murder in the first degree. Rush was likewise a principal founder, in 1786, of the Philadelphia Dispensary for the Poor, the first free dispensary in the United States.

As a physician, Rush, like so many of his medical contemporaries, is a characteristic figure of the transitional period during which modern medicine was just beginning to emerge. Medical practice was based on empiricism, tradition, and theories that lacked any firm scientific foundation. Moreover, like other physicians of his time, Rush supported his opinion and practice by argument rather than experiment.

Rush's most significant contribution to medicine and to social welfare was his study of mental illness. In 1812 he published *Medical Inquiries and Observations Upon the Diseases of the Mind*, the first American treatise on mental illness. It remained the only American work of its kind for more than fifty years after its publication. Basing his treatise on observations made at the Pennsylvania Hospital, where he was in charge of mental patients for some thirty years, Rush attempted to provide a systematic account of mental disorders in terms of theory and practice. He viewed mental derangement as basically due to disease of the cerebral arteries, and his therapy was intended to relieve this pathology. Treatment was designed to affect the circulation of the blood and included purges, emetics, bleeding, a bland diet, and hot

and cold showers. At the same time, Rush was aware of psychic components in mental illness and urged the patients to write down what troubled their minds and to read what they had written. It is clear that he was aware of the ill effects of repressed emotions. Rush was humane to his mental patients, but he recommended coercion under some circumstances and even advocated corporal punishment in extreme situations. Like his contemporary Philippe Pinel, Rush knew the importance of enlightened hospital management in dealing with mental patients. It was due to his efforts that a separate wing for the insane was added at the Pennsylvania Hospital in 1796. Rush likewise recognized the need for well-qualified persons to serve as attendants in establishments for the mentally ill, as well as the necessity of keeping the patients and their quarters as neat and clean as possible. He also advocated occupational therapy; the separation of those patients who were disoriented and violent from the others; the provision of exercise and recreation for the patients; and the treating of the mentally ill with respect and kindness.

Rush exerted a significant influence on the practitioners of his time and for several generations after his death. His most important achievement was his pioneer endeavor to raise the study and treatment of the mentally ill to a systematic, scientific level.

GEORGE ROSEN

[See also PSYCHIATRY, article on CHILD PSYCHIATRY; and the biography of PINEL.]

#### WORKS BY RUSH

1768 *De coctione ciborum in ventriculo*. Edinburgh: Balfour.

- (1786) 1839 *Inquiry Into the Influence of Physical Causes Upon the Moral Faculty*. Philadelphia: Cist. → Delivered before the American Philosophical Society, February 27, 1786.  
 (1812) 1930 *Medical Inquiries and Observations Upon the Diseases of the Mind*. 4th ed. Philadelphia: Grigg.  
*The Autobiography of Benjamin Rush*. Edited with an introduction and notes by George W. Corner. Princeton Univ. Press, 1948.  
*The Letters of Benjamin Rush*. Edited by L. H. Butterfield. 2 vols. Princeton Univ. Press, 1951.  
*The Selected Writings of Benjamin Rush*. Edited by Dagobert D. Runes. New York: Philosophical Library, 1947.

#### SUPPLEMENTARY BIBLIOGRAPHY

- BINGER, CARL 1966 *Revolutionary Doctor: Benjamin Rush*. New York: Norton.  
 DAIN, NORMAN 1964 *Concepts of Insanity in the United States: 1789-1865*. New Brunswick, N.J.: Rutgers Univ. Press. → See especially pages 15-25.  
 DEUTSCH, ALBERT (1937) 1949 Benjamin Rush: The Father of American Psychiatry. Pages 72-87 in Albert Deutsch, *The Mentally Ill in America*. Rev. ed. New York: Columbia Univ. Press.  
 GOODMAN, NATHAN C. 1934 *Benjamin Rush: Physician and Citizen, 1746-1813*. Philadelphia: Univ. of Pennsylvania Press.  
 LLOYD, JAMES H. 1930 Benjamin Rush and His Critics. *Annals of Medical History* 2:470-475.  
 PEPPER, O. H. PERRY 1946 Benjamin Rush's Theories on Bloodletting After One Hundred and Fifty Years. College of Physicians of Philadelphia, *Transactions and Studies* Fourth Series 14:121-126.  
 ROSEN, GEORGE 1952 Political Order and Human Health in Jeffersonian Thought. *Bulletin of the History of Medicine* 26:32-44.  
 SHRYOCK, RICHARD H. 1945 The Psychiatry of Benjamin Rush. *American Journal of Psychiatry* 101:429-432.  
 SHRYOCK, RICHARD H. 1946 Benjamin Rush From the Perspective of the Twentieth Century. College of Physicians of Philadelphia, *Transactions and Studies* Fourth Series 14:113-120.  
 WITTELS, FRITZ 1946 Contribution of Benjamin Rush to Psychiatry. *Bulletin of the History of Medicine* 20:157-166.





# S

## SACRED AND PROFANE

See POLLUTION; RELIGION; and the biographies of BECKER and GENNEP.

## SAINT-SIMON

Claude Henri de Rouvroy, Comte de Saint-Simon (1760–1825), French social philosopher and reformer, is a controversial figure in modern social thought, who—without writing a single enduring work—had a crucial role in the early nineteenth-century developments of industrial socialism, positivism, sociology, political economics, and the philosophy of history. In his final years he inspired a Christian socialist movement with a secular gospel of human brotherhood. By scholarly standards his works—mostly reformist essays and brochures—are highly deficient: overly polemical, weak in organization and conceptual clarity, and often marred by confusions between questions of value and fact. Much of his subsequent influence is due to students and disciples, who systematized, popularized, and also partly reshaped his ideas, and especially to Auguste Comte, his onetime protégé and collaborator, whose work in sociology and philosophy greatly overshadows Saint-Simon's.

In retrospect, Saint-Simon's main significance for the social sciences is threefold. He was one of the first to grasp the revolutionary implications of "industrialization" (a word he himself coined) for traditional institutions and morality and to conceptualize the industrial system as a distinctive type. He was also among the earliest to advocate a naturalistic science of society as a rational guide to social reconstruction. But he is most important as provisional formulator of an "evolutionary organ-

icist" theory, whose influence is reflected in social evolutionary doctrines as diverse as those of Herbert Spencer, Lester Ward, and Karl Marx. He directly inaugurated the "positivist organicist" school—most notably represented by Comte and Émile Durkheim—which for a century thereafter was to vie with utilitarianism, Marxism, and Hegelian historicism for theoretical predominance in the social sciences (Martindale 1960, part 2). Through Durkheim, his organicist concept of social order carries over into contemporary "functionalism" in anthropology and sociology.

Recent biographers have depicted Saint-Simon as a brilliant, but erratic, figure, possessed from early life with a consuming passion for greatness which led him into various endeavors, turning to serious studies and his writings after the most varied social experiences. He was born in Paris to the socially marginal, relatively impoverished side of a prominent noble family that claimed descent from Charlemagne. His boyhood education, through tutoring, included conventional exposure to Enlightenment ideas and Catholicism, both profoundly influential in his later thought. At 17 he was commissioned into the army, remaining for four years, during which he fought with some distinction in the American Revolution. From this time until the French Revolution he tried a variety of ambitious promotional ventures with little success. During the Terror of 1793–1794 he was imprisoned for a year, narrowly escaping a death sentence. This experience left him with a lasting dread of revolutionary violence. After his release, he amassed a considerable fortune—by speculating in confiscated church lands—which he lavished on an extravagant Paris salon, catering particularly to scientists and intellectuals, from whom he ac-



quired a broad, if superficial, education. But his funds were spent by 1804, and he lived his last twenty years in modest circumstances, at times verging on poverty, largely cut off from his former glittering life.

### Social theory

At 42, with poverty threatening, Saint-Simon took up his new career as writer and social reformer. Throughout the turbulence of the Napoleonic empire and the Bourbon restoration, he presented a many-sided analysis of problems confronting Europe, proposing numerous remedies that gradually broadened into a full-scale social reconstruction program. The emphasis shifts at several points, and few of the ideas, taken singly, are new. What is new is the underlying rationale of evolutionary organicism: a partly conscious synthesis of the scientific naturalism and rationalist faith of Condorcet and the Enlightenment, the more practical materialism of the rising entrepreneurial bourgeoisie; and the authoritarian emphasis upon ideological unity found in Burke and in restorationists like Bonald and de Maistre.

The starting point is the conservative ideal of "organic" equilibrium, according to which society is unified under a single organizational system, undisturbed by conflicts between individuals or social classes. To Saint-Simon such integration requires both a "temporal" system of institutions (for government, the economy, and other coordinative activities) and a "spiritual" ideology of worldly existence and conduct that would give coherent justification to the temporal system and command the allegiance of the highly educated as well as the general populace. Under such an established order, individuals are viewed as beholden to the system for survival and well-being, and a mutual interest of all social classes in the system's success is assumed.

Conservative versions of organicism had generally accepted a single equilibrium state with traditional sanction. Saint-Simon was among the first to relativize organicism by adding the evolutionary idea of social development. He conceived of an orderly progression of stable organic civilizations that represent different stages of advancement. Each system in turn is viewed as appropriate to its time of formation, only to become obsolete as a higher level is reached. Thus, in the history of Western man, Saint-Simon discerned two main organic systems. The first, predominant in ancient Greece and Rome, had a polytheistic ideology, a slave economy, and monolithic political rule. By

the eleventh century, it had been supplanted by the second, characterized by feudalism and Catholicism, after centuries of disorganization and conflict between the two rival systems. The triumph of medievalism was assured by its superiority over its predecessor in promoting a more rational theology, a more humane form of bondage, the separation of temporal from spiritual power, and technological improvements. In this analysis, Saint-Simon, like the Enlightenment philosophers, accepted the growth of knowledge and ethical understanding as primary criteria and determinants of evolutionary progress. He posited a broad, overriding trend in human history, from anthropomorphically theistic interpretations of the world to scientific naturalism, and, correlatively, from domination by the strong to productive cooperation and humane morality. As enlightenment progresses, new civilizations are formed, based on new ideologies initially introduced by advanced elites and subsequently reflected in popular institutions. But the transition from one system to another is very slow, since organic civilizations require centuries to develop and, once established, are strongly resistant to change.

Saint-Simon believed that all Europe had been in a transitional crisis since the fifteenth century, when medievalism began to give way to a new system founded on industry and science. He wrote as the new system's advocate, urging heads of state and influential groups to hasten its inception as the only way to restore stability; he was, thus, the first of many meliorists to present reforms as evolutionary necessity ("Our intention merely is to promote and explain the inevitable."). His early writings during Napoleon's reign stress ideological aspects of the crisis and contain his main positivist doctrines (1807; 1813a; 1813b). The groundwork of positivism as an empiricist philosophy of knowledge had already been established by Bacon, Hume, and others. Saint-Simon's contribution, later amplified in Comte's more rigorous work, was its evolutionary reformulation as part of a comprehensive reconstruction program. He argued that theistic Roman Catholicism—the spiritual basis of medievalism—had been undermined by the sciences and by such secularists as Voltaire and the Encyclopedists. But these modern intellectual forces, while aiding progress by exposing an outdated system, had been negativistic in failing to provide an adequate replacement. Needed, instead, was a new, encompassing "positive" philosophy based on the sciences, as fundament for the new social order. All sciences were to contribute, including a new

social science (called social physiology, to indicate that social phenomena were to be viewed as part of life's natural development) that would "inductively" derive the new moral codes and policies for the future society by historical extrapolations—a task for which no explicit methodology was offered. A new scientific advisory group would insure the application of results to social policy, so "that men henceforth [would] do consciously and more effectively what they previously [had] done slowly, unconsciously and indecisively" (*Oeuvres de Saint-Simon*, vol. 18, p. 166). Popular attitudes were to be correspondingly changed by a scientific program of mass education that would eventually establish a nontheistic religion of "physicism." Thus, Saint-Simon envisioned nothing less than a total scientific transformation of Western civilization and implied that while positivism would serve as a new philosophy for the educated, it would function as a religion for the masses.

After Napoleon's downfall Saint-Simon shifted his attention from the ideology of the new system to its temporal structure and policies. In a series of periodicals, *L'industrie* (1816–1818), *Le politique* (1819), *L'organisateur* (1819–1820), and *Du système industriel* (1821–1822), he supported the third estate against the re-established nobles and clergy. These periodicals contained many of his own articles presenting his main socialist doctrines—but a strange socialism it is, resembling venture capitalism and technocracy as much as Marxism or primitive communism. In contrast with the latter, he abandoned the ideal of the small equalitarian communal unit, still partly upheld by Fourier and others. The future society portrayed is, above all, one of productive achievement, with poverty and war eradicated through bold, large-scale industrialization under planned scientific guidance. It is an open-class system in which caste privileges are abolished, work is provided for all, and rewards are assigned according to merit. The state changes from "government" (characterized by class domination and national rivalries) to a welfare system managed scientifically by expert public servants who are mainly concerned with economic regulation; the power of the state is drastically reduced, and what is left is transferred from war lords and idlers to the productive workers. But, unlike Marx, Saint-Simon opposed all violence, whether revolutionary or not, hoping that the feudal rulers would relinquish their power peacefully as they became persuaded of their fate. He also differed from Marx in viewing bourgeoisie, scientists, and proletariat alike as members of the

productive working class: all are natural allies in the struggle against feudalism and cobeneficiaries of the future industrial system. In fact, he saw bankers, engineers, and manufacturers (under scientific advisement) as the best qualified revolutionary leaders of the working-class coalition and hoped that, through their managerial effectiveness and wisdom, destructive class struggles would disappear in the future. The most basic contrast, however, is that despite Saint-Simon's emphasis on material prosperity as a political goal, he was more an ideological than an economic determinist. The desired evolution ultimately depends, for him, on the enlightened good will of those in power, and to foster this good will, in his final work, "New Christianity" (1825), he espoused a secular Christian revival that would instill in all social classes an awareness of their common destiny.

### Influence

Saint-Simon's ideas caused little stir until his death, when a movement was formed in his name, inspired by "New Christianity" and propounding a gospel of brotherly love, concern for the poor, and the reconciliation of spiritual values with material progress. During eight stormy years of organized existence, it attracted surprisingly many of Europe's outstanding young intellectuals (at various times including Carlyle, Mill, Buchez, Carnot, Blanc, Chevalier, and Heine, to name only a few) by appealing to a rising discontent with capitalism and a sense that reform was needed. First led by Bazard, the movement reworked Saint-Simon's ideas into the famous *Doctrine of Saint-Simon* (1830; often dubbed socialism's "Old Testament"), for which he is best remembered. But while the doctrine retained and made more coherent Saint-Simon's evolutionary formula, concept of crisis, stress on planning, pacifism, work ethic, and call for restored social unity, it also introduced important changes. In a position closer to Marxism, it condemned capitalist exploitation along with feudalism; abolition of inheritance, socialization of property, and equality of the sexes were all added to the program; the whole approach became more authoritarian and antiscientific in spirit. This trend came to a climax in 1832, after a bitter schism in which, under Enfantin's charismatic direction, one faction of Saint-Simonians became a hierarchical religious sect, complete with mystical liturgy, substituting a sensate monism of flesh and spirit for the remnants of scientific empiricism. Enfantin quickly fell into disrepute for attacking marital "hypocrisy" and advocating "free love" (which



Saint-Simon himself had indulged in but never preached), and the sect was disbanded for its heresies; but not before its earlier doctrine had become a part of the intellectual heritage of the West.

MARTIN U. MARTEL

[See also ELITES; EVOLUTION, article on SOCIAL EVOLUTION; SOCIALISM; and the biographies of COMTE; CONDORCET; DURKHEIM; FOURIER; LE PLAY; MARX; OWEN; SPENCER; WARD, LESTER F.]

#### WORKS BY SAINT-SIMON

The periodicals edited by Saint-Simon, *L'Industrie* (1816-1818), *Le politique* (1819), *L'organisateur* (1819-1820), and *Du système industriel* (1821-1822), are often catalogued by libraries under Saint-Simon's name. Manuel 1962 describes them as appearing intermittently, in part to evade the rules of censorship applied to serial publications, but chiefly because Saint-Simon found it difficult to raise the money necessary to publish them.

1807 *Introduction aux travaux scientifiques du dix-neuvième siècle*. Paris: Scherff.

(1813a) 1876 *Mémoire sur la science de l'homme*. Volume 40 of *Oeuvres de Saint-Simon et d'Enfantin*. Paris: Dentu.

1813b *Travail sur la gravitation universelle*. Paris. → No publisher given.

(1825) 1952 *New Christianity: Dialogue*. Pages 81-116 in Saint-Simon, *Selected Writings*. Edited and translated by F. M. H. Markham. Oxford: Blackwell. → First published in French. A new French edition was published in 1943 by Aubry.

*Henri de Saint-Simon: Social Organization*. New York: Harper, 1964. → Also published in 1952 by Macmillan under the title *Henri de Saint-Simon: Selected Writings*.

*Oeuvres de Claude-Henri de Saint-Simon*. 6 vols. Paris: Éditions Anthropos, 1966. → Volumes 1-5 reprinted from *Oeuvres de Saint-Simon et d'Enfantin*, 1865-1878. Volume 6 reprinted from other works.

*Oeuvres de Saint-Simon et d'Enfantin*. 47 vols. Paris: Dentu, 1865-1878. → Saint-Simon's writings are in Volumes 15, 18-23, and 37-40.

*Selected Writings*. Edited and translated with an introduction by F. M. H. Markham. Oxford: Blackwell, 1952.

#### SUPPLEMENTARY BIBLIOGRAPHY

CHARLÉTY, SEBASTIEN (1896) 1931 *Histoire du Saint-Simonisme (1825-1864)*. New ed. Paris: Hartmann.

COLE, G. D. H. 1953-1960 *A History of Socialist Thought*. 5 vols. New York: St. Martins; London: Macmillan.

*The Doctrine of Saint-Simon; an Exposition: First Year, 1828-1829*. Translated by Georg C. Iggers. (1830) 1958 Boston: Beacon. → First published in French. The translation is based on the third, revised and enlarged edition.

DURKHEIM, ÉMILE (1928) 1958 *Socialism and Saint-Simon (Le socialisme)*. Yellow Springs, Ohio: Antioch. → First published in French. A paperback edition was published in 1962 by Collier.

GOUIER, HENRI G. 1933-1941 *La jeunesse d'Auguste Comte et la formation du positivisme*. 3 vols. Paris: Vrin.

HALÉVY, ÉLIE (1938) 1965 *The Era of Tyrannies: Essays on Socialism and War*. Translated by R. K. Webb. New York: Doubleday. → First published in French.

HAYEK, FRIEDRICH A. VON 1952 *The Counter-revolution of Science: Studies on the Abuse of Reason*. Glencoe, Ill.: Free Press.

MANUEL, FRANK E. 1956 *The New World of Saint-Simon*. Cambridge, Mass.: Harvard Univ. Press.

MANUEL, FRANK E. 1962 *The Prophets of Paris*. Cambridge, Mass.: Harvard Univ. Press. → A paperback edition was published by Harper in 1965.

MARTINDALE, DON 1960 *The Nature and Types of Sociological Theory*. Boston: Houghton Mifflin.

## SALES TAXES

See under TAXATION.

## SAMPLE SURVEYS

There is hardly any part of statistics that does not interact in some way with the theory or the practice of sample surveys. The differences between the study of sample surveys and the study of other statistical topics are primarily matters of emphasis.

The field of survey research is closely related to the statistical study of sample surveys [see SURVEY ANALYSIS]. Survey research is more concerned with highly multivariate data and complex measures of relationship; the study of sample surveys has emphasized sampling distributions and efficient design of surveys.

### I. THE FIELD

W. Edwards Deming

### II. NONPROBABILITY SAMPLING

Alan Stuart

## I

### THE FIELD

The theory of sample surveys is mathematical and constitutes a part of theoretical statistics. The practice of sample surveys, however, involves an intimate mixture of subject matter (such as demography, psychology, consumer research, medicine, engineering) with theory. The germ of a study lies in the subject matter. Translation of a substantive question into a stimulus (question or test) enables man to inquire of nature and to quantify the result in terms of estimates of what the same inquiry would produce were it to cover every unit of the population.

Sampling, properly applied, does more. It furnishes, along with an estimate, an index of the precision thereof—that is, a margin of the uncertainty, for a stated probability, that one may reasonably ascribe to accidental variations of all kinds, such as variability between units (that is,

between households, blocks, patients), variability of the interviewer or test from day to day or hour to hour, variations in coding, and small, independent, accidental errors in transcription and card punching.

The techniques of sampling also enable one to test the performance of the questionnaire and of the investigators and to test for differences between alternative forms of the questionnaire. They enable one to measure the extent of under-coverage or over-coverage of the prescribed units selected and also to measure the possible effects of differences between investigators and of departures from prescribed rules of interviewing and coding.

This article describes *probability sampling*, with special reference to studies of human populations, although the same theory and methods apply to studies of physical materials, to accounting, and to a variety of other fields. The main characteristic of probability sampling is its use of the theory of probability to maximize the yield of information for an allowable expenditure of skills and funds. Moreover, as noted above, the same theory enables one to estimate, from the results themselves, margins of uncertainty that may reasonably be attributed to small, accidental, independent sources of variation. The theory and practice of probability sampling are closely allied to the design of experiments.

The principal alternatives to probability sampling are judgment sampling and convenience sampling [see SAMPLE SURVEYS, *article on NON-PROBABILITY SAMPLING*].

**Uses of sampling.** Probability sampling is used in a wide variety of studies of many different kinds of populations. Governments collect and publish monthly or quarterly current information in such areas as employment, unemployment, expenditures and prices paid by families for food and other necessities, and condition and yield of crops.

In modern censuses only basic questions are asked of every person, and most census information is elicited for only a sample of the people, such as every fourth household or every twentieth. Moreover, a large part of the tabulation program is carried out only on a sample of the information elicited from everyone.

Sampling is the chief tool in consumer research. Samples of records, often supplemented by other information, furnish a basis on which to predict the impact that changes in economic conditions and changes in competitive products will have on a business.

Sampling is an important tool in supervision and

is helpful in many other administrative areas, such as studies of use of books in a library to improve service and to make the best use of facilities.

**Sampling—what is it?** Everyone acquires information almost daily from incomplete evidence. One decides on the basis of the top layer of apples in a container at the fruit vendor's whether to buy the whole container. The top layer is a good sample of the whole if the apples are pretty well mixed; it is a bad sample and may lead to a regrettable purchase if the grocer has put the best ones on top.

The statistician engaged in probability sampling takes no chances on inferences drawn exclusively from the top layer or from any other single layer. He uses random numbers to achieve a standard degree of mixing, thereby dispersing the sample throughout the container and giving to every sampling unit in the frame an ascertainable probability of selection [see RANDOM NUMBERS]. He may use powerful techniques of stratification, ratio estimation, etc., to increase accuracy and to decrease costs. For instance, in one type of stratified sampling he in effect divides the container of apples into layers, mixes the apples in each layer, and then takes a sample from each layer.

**Some history of sampling.** Sir Frederick Morton Eden estimated the number of inhabitants of Great Britain in 1800 at nine million, using data on the average number of people per house in selected districts and on the total number of houses on the tax-rolls, with an allowance for houses not reported for taxation. The first census of Great Britain, in 1801, confirmed his estimate. Messance in 1765 and Moheau in 1778 obtained estimates of the population of France by multiplying the ratio of the number of inhabitants in a sample of districts to the number of births and deaths therein by the number of births and deaths reported for the whole country. Laplace introduced refinements in 1786 and calculated that 500,000 was the outside margin of error in his estimate of the population of France, with odds of 1,161 : 1. His estimate and its precision were more successful than those of the complete census of France that was attempted at the same time. [See LAPLACE.]

A. N. Kiaer used systematic selection in a survey of Norwegian workers in 1895, as well as in special tabulations from the census of Norway in 1900 and from the census of Denmark in 1901 and in a study of housing in Oslo in 1913.

Bowley in 1913 used a systematic selection of every twentieth household of working-class people in Reading (England) and computed standard errors of the results.



Tabulation of the census of Japan in 1921, brought to a halt by the earthquake of 1923, went forward with a sample consisting of the records of every thousandth household. The results agreed with the full tabulation, which was carried out much later. The Swedish extraordinary census of 1935 provides a good example of the use of sampling in connection with total registrations.

One strong influence on American practice came in the 1930s from Margaret H. Hogg, who had worked under Bowley. Another came when controversies over the amount of unemployment during the depressions of 1921 and 1929 called for improved methods of study—Hansen's sample of postal routes for estimates of the amount of unemployment in 1936 gained recognition for improved methods; without it the attempt at complete registration of unemployed in the United States at the same time would have been useless.

Mahalanobis commenced in 1932 to measure the yield of jute in Bengal and soon extended his surveys to yields of rice and of other crops. In 1952 all of India came under the national surveys, the scope of which included social studies and studies of family budgets, sickness, births, and deaths. Meanwhile, the efforts of statisticians, mainly in India and England, had brought advances in methodology for estimation of yield per acre by random selection of small plots to be cut and harvested.

A quarterly survey of unemployment in the United States, conducted through interviews in a sample of households within a sample of counties, was begun in 1937. It was soon made monthly, and in 1942 it was remodeled much along its present lines (Hansen et al. 1953, vol. 1, chapter 9).

Sampling was used in the census of the United States in 1940 to extend coverage and to broaden the program of tabulation and publication. Tabulation of the census of India in 1941 was carried out by a 2 per cent sample. Subsequent censuses in various parts of the world have placed even greater dependence on sampling, not only for speed and economy in collection and tabulation but also for improved reliability. The census of France used sampling as a control to determine whether the complete Census of Commerce of 1946 was sufficiently reliable to warrant publication; the decision was negative (Chevry 1949). [For further history, see Stephan 1948. Some special references to history are contained in Deming (1943) 1964, p. 142. See also STATISTICS, article on THE HISTORY OF STATISTICAL METHOD.]

**Misconceptions about sampling.** Sampling, of course, possesses some disadvantages: it does not

furnish detailed information concerning every individual person, account, or firm; furthermore, error of sampling in very small areas and sub-classes may be large. Many doubts about the value of sampling, however, are based on misconceptions. Some of the more common misconceptions will now be listed and their fallacies pointed out.

*It is ridiculous to think that one can determine anything about a population of 180 million people, or even 1 million people, from a sample of a few thousand.* The number of people in the country bears almost no relation to the size of the sample required to reach a prescribed precision. As an analogy (suggested by Tukey), consider a basket of black and white beans. If the beans are really mixed, a cupful would determine pretty accurately the proportion of beans that are black. The cupful would still suffice and would give the same precision for a whole carload of beans, provided the beans in the carload were thoroughly mixed. The problem lies in mixing the beans. As has already been noted, the statistician accomplishes mixing by the use of random numbers.

*Errors of sampling are a hazard because they are ungovernable and unknown. Reliability of a sample is a matter of luck.* Quality and reliability of data are built in through proper design and supervision, with aid from the theory of probability. Uncertainty resulting from small, independent, accidental errors of a canceling nature and variation resulting from the use of sampling are in any case determinable afterward from the results themselves.

*Errors of sampling are the only danger that one has to worry about in data.* Uncertainty in statistical studies may arise from many sources. Sampling is but one source of error. [See below, and see also ERRORS, article on NONSAMPLING ERRORS].

*Electronic data-processing machines, able to store and retrieve information on millions of items with great speed, eliminate any need of sampling.* This is a fanciful hope. The inherent accuracy of original records as edited and coded is the limitation to the accuracy that a machine can turn out. Often, complete records are flagrantly in error or fail to contain the information that is needed. Moreover, machine-time is expensive; sampling reduces cost by reducing machine-time.

*A "complete" study is more reliable than a sample.* Data are the end product of preparation and of a long series of procedures—interviewing, coding, editing, punching, tabulation. Thus, error of sampling is but one source of uncertainty. Poor workmanship and structural limitations in the method of test or in the questionnaire affect a

complete count as much as they do a sample. It is often preferable to use funds for improving the questionnaire and tests rather than for increasing the size of the sample.

**Statistical parts of sampling procedure.** A sampling procedure consists of ten parts. In the following list, *M* will denote those parts that are the responsibility of the expert on the subject matter, and *S* will denote those that are the responsibility of the statistician. (The technical terms used will be defined below.)

(a) Formulation of the problem in statistical terms (probability model) so that data will be meaningful (*M, S*). A problem is generated by the subject matter, not by statistical theory.

(b) Decision on the universe (*M*). The universe follows at once from a careful statement of the problem.

(c) Decision on the frame (*M, S*). Decision on the type and size of sampling units that constitute the frame (*S*).

(d) Procedure for the selection of the sample (*S*).

(e) Procedure for the calculation of estimates of the characteristics desired (averages, totals, proportions, etc.) (*S*).

(f) Procedure for the calculation of standard errors (*S*).

(g) Design of statistical controls, to permit detection of the existence and extent of various non-sampling errors (*S*).

(h) Editing, coding, tabulation (*M, S*).

(i) Evaluation of the statistical reliability of the results (*S*).

(j) Uses of the data (*M*).

### Definitions of terms

The technical terms that have been used above and that will be needed for further discussion will now be defined.

**Universe of study.** The universe consists of all the people, firms, material, conditions, units, etc., that one wishes to study, whether accessible or not. The universe for any study becomes clear from a careful statement of the problem and of the uses intended for the data. Tabulation plans disclose the content of the universe and of the information desired. Examples of universes are (i) the housewives aged 20–29 that will live in the Metropolitan Area of Detroit next year, (ii) all the school children in a defined area, (iii) all the pigs in a country, both in rural areas and in towns.

**Frame.** The frame is a means of access to the universe (Stephan 1936) or to enough of the universe to make a study worthwhile. A frame is composed of sampling units. A sampling unit

commonly used in house-to-house interviewing is a compact group or segment of perhaps five consecutive housing units. A frame is often a map, divided up—either explicitly or tacitly—into labeled areas. In a study concerned with professional men, for example, the frame might be the roster of membership of a professional society, with pages and lines numbered. The sampling unit might be one line on the roster or five consecutive lines.

Without a frame probability sampling encounters numerous operational difficulties and inflated variances (see, for example, the section "Sampling moving populations," below).

In the types of problems to be considered here (with the exception of those treated in the section "Sampling moving populations," below) there will be a frame, and every person, or every housing unit, will belong to one sampling unit, or will have an ascertainable probability of belonging to it. In the sampling of stationary populations, a sampling procedure prescribes rules by which it is possible to give a serial number to any sampling unit, such as a small area. A random number will then select a definite sampling unit from the frame and will lead to investigation of all or a subsample of the material therein that belongs to the universe.

**Selection of persons within a dwelling unit.** Some surveys require information concerning individuals, and in such cases it may be desirable, for various reasons (contagion, fatigue, and so on), to interview only one eligible person in a dwelling unit that lies in a selected segment. In such surveys, the interviewer may make a list of the eligible people in each dwelling unit that falls in the sample and may select therefrom, on the spot, by a scheme based on random numbers, one person to interview. Appropriate weights are applied in tabulation (Deming 1960, p. 240).

**Nominal frame and actual frame.** One must often work with a frame that fails to include certain areas or classes that belong to the universe. A list of areas that contain normal families may not lead to all the consumers of a product, as some consumers may live in quasi-normal quarters, such as trailers and dormitories. Extension of the sampling procedure into these quarters may present problems. Fortunately, the proportion of people in quasi-normal households is usually small (mostly 1 per cent to 3 per cent in American cities), and one may therefore elect to omit them.

A frame may be seriously impaired if it omits too much of certain important classes that by definition belong to the nominal frame. It is substantive judgment, aided by calculation, that must decide whether a proposed frame is satisfactory.



*Sampling from an incomplete frame.* Almost every frame is in some respects out of date at the time of use. It is often possible, however, to use an obsolete or incomplete frame in a way that will erase the defects in the areas that fall into the sample. One may, for example, construct rules by which to select large sampling units from an incomplete frame and then to amend those units, by local inquiry, in order to bring them up to date. Selection of small areas within the larger area, with the appropriate probability, will maintain the prescribed over-all probability of selection.

*Sampling for rare characteristics.* One sometimes wishes to study a rare class of people when there is no reliable list of that class. One way to accomplish this is to carry out a cheap, rapid test in order to separate a sample of households into two groups (strata)—one group almost free of the rare characteristic, the other heavily populated with it—and then to investigate a sample drawn from each group. Optimum sampling fractions and weights for consolidation may be calculated by the theory of stratified sampling (discussed below; see also Kish in Symposium . . ., 1965).

*Equal complete coverage of a frame.* The equal complete coverage of a frame is by definition the result that would be obtained from an investigation of all sampling units in a given frame, carried out by the same field workers or inspectors, using the same definitions and procedures, and exercising the same care as they exercised on the sample, and at about the same period of time. The adjective "equal" signifies that the same methods must be used for the equal complete coverage as for the sample.

*Some operational definitions. Sampling error.* Suppose that for a given frame, sampling units bear the serial numbers 1, 2, 3, and on to  $N$ . However it be carried out, and whatever be the rules for coding and for adjustment for nonresponse, a complete coverage of the  $N$  sampling units of the frame would yield the numerical values

$$a_1, a_2, a_3, \dots, a_N \text{ for } x, \\ b_1, b_2, b_3, \dots, b_N \text{ for } y.$$

In a survey of unemployment, for example, the  $x$ -characteristic of a person might be the property of being unemployed and his  $y$ -characteristic the property of belonging to the labor force. Then  $a_1$ , the  $x$ -population of sampling unit No. 1 (which might consist of five successive households), would be the count of people that have the  $x$ -characteristic in that sampling unit. That is,  $a_1$  would be the count of unemployed persons in the five households. Similarly,  $b_1$ , the  $y$ -population, would be the count of people in the labor force in those same

households. Then  $a_1/b_1$  would be the proportion unemployed in the sampling unit of five households.

Again,  $x$  might refer to expenditure for bread and  $y$  to expenditure for all food. Then  $a_1/b_1$  would be the proportion of money that goes for bread in sampling unit No. 1, expenditure for all food being the base.

Here, the people with the  $x$ -characteristic form a subclass of those with the  $y$ -characteristic, but this may not be so in other surveys. Thus, the  $x$ -characteristic and the  $y$ -characteristic might form a dichotomy, such as passed and rejected or male and female. One often deals with multiple characteristics, but two will suffice here.

Denote the sum of the  $x$ -values and of the  $y$ -values in the  $N$  sampling units by

$$A = a_1 + a_2 + a_3 + \dots + a_N = Na = x\text{-total}, \\ B = b_1 + b_2 + b_3 + \dots + b_N = Nb = y\text{-total},$$

which makes  $a$  and  $b$  the average  $x$ -value and the average  $y$ -value per sampling unit in the frame, as in Table 1. For example,  $A$  might be the total number unemployed in the whole frame and  $B$  the total number of people in the labor force. Then  $\phi = A/B$  would be the proportion of people in the labor force that are unemployed.

An operational definition of the sampling process and of the consequent error of sampling is contained in the following experiment.

(a) Take for the frame  $N$  cards, numbered serially 1 to  $N$ . Card  $i$  shows  $a_i$  and  $b_i$  for the values of the  $x$ -characteristic and  $y$ -characteristic.

(b) Draw a sample of  $n$  cards, following the specified sampling procedure (which will invariably require selection by random numbers).

Table 1 illustrates the notation for the frame and for the results of a sample. The serial numbers on the cards in the sample are not their serial numbers in the frame but denote instead the ordinal number as drawn by random numbers. Sample card No. 1 could be any card from 1 to  $N$  in the frame. In general, another sample would be composed of different cards, as the drawings are random.

(c) Form estimates by the formulas specified in the sampling plan. For illustration, one may form, from the sample, estimators like

$$(1) \quad \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i,$$

$$(2) \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i,$$

$$(3) \quad X = N\bar{x},$$

$$(4) \quad Y = N\bar{y},$$

$$(5) \quad f = \bar{x}/\bar{y}.$$

Table 1 — Some notation for frame and sample

FRAME			SAMPLE		
Serial number of sampling unit	x-value	y-value	Serial number in order drawn in sample	x-value	y-value
1	$a_1$	$b_1$	1	$x_1$	$y_1$
2	$a_2$	$b_2$	2	$x_2$	$y_2$
⋮	⋮	⋮	⋮	⋮	⋮
$N$	$a_N$	$b_N$	$n$	$x_n$	$y_n$
Total	$A$	$B$	$\bar{x}$	$\bar{y}$	
Average per sampling unit	$a = A/N$	$b = B/N$	$\bar{x} = x/n$	$\bar{y} = y/n$	
Variance*	$\sigma_a^2 = \frac{1}{N} \sum_{i=1}^N (a_i - a)^2$	$\sigma_b^2 = \frac{1}{N} \sum_{i=1}^N (b_i - b)^2$	$s_x^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$	$s_y^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2$	
Standard deviation	$\sigma_a$	$\sigma_b$	$s_x$	$s_y$	

\* Some authors define variances by means of  $N-1$  and  $n-1$  rather than  $N$  and  $n$ .

If (1), (2), (3), (4), and (5) are used as estimators of  $a$ ,  $b$ ,  $A$ ,  $B$ , and  $\phi$ , respectively, and if the results of the complete coverage were known, then one could, for any experiment, compute errors of sampling, such as

$$\begin{aligned} (6) \quad & \Delta\bar{x} = \bar{x} - a, \\ (7) \quad & \Delta\bar{y} = \bar{y} - b, \\ (8) \quad & \Delta f = f - \phi. \end{aligned}$$

It is an exciting fact that a single sample—provided that it is big enough (usually 25, 30, or more sampling units), and provided that it is designed properly and skillfully in view of possible statistical peculiarities of the frame and is carried out in reasonable conformance with the specifications—will make possible an estimate, based on theory to follow later, of the important characteristics of the distribution of sampling variation of all possible samples that could be drawn from the given equal complete coverage and processed by the specified sampling procedure.

**Standard error and mathematical bias.** We continue our conceptual experiment.

(d) Return the sample of  $n$  cards to the frame, and repeat steps (b) and (c) by the same sampling procedure, to form a new sample and new estimates  $\bar{x}$ ,  $\bar{y}$ ,  $f$ . Repeat these steps again and again, 10,000 times or more.

Explicit statements will now be confined to  $\bar{x}$ . The 10,000 experiments give an empirical distribution for  $\bar{x}$ , by which one may count the number of samples for which  $\bar{x}$  lies between, for example, 100 and 109. We visualize an underlying theoretical distribution of  $\bar{x}$ , which the empirical dis-

tribution approaches closer and closer as the number of repetitions increases.

We are typically concerned with relationships between (i) the empirical distribution of  $\bar{x}$  and (ii) the theoretical distribution of  $\bar{x}$ , for the given sampling procedure. Study of these relationships helps in the use of sampling for purposes of making estimates of characteristics of the frame.

Let  $a$  be the characteristic of the complete coverage that the generic symbol  $\bar{x}$  estimates. Then if

$$(9) \quad E\bar{x} = a,$$

the sampling procedure is said to be unbiased. (The symbol  $E$  denotes *expectation*, the mean of the theoretical distribution of  $\bar{x}$ .) But if

$$(10) \quad E\bar{x} = a + C, \quad C \neq 0,$$

the sampling procedure has the mathematical bias  $C$ . In any case, the variance of the distribution of  $\bar{x}$  is

$$(11) \quad \sigma_{\bar{x}}^2 = E(\bar{x} - E\bar{x})^2,$$

and its square root,  $\sigma_{\bar{x}}$ , is the standard error of the sampling procedure for the estimator  $\bar{x}$ . Thus, a sampling procedure has, for any estimator, an expected value, a standard error, and possibly a mathematical bias (see the section "Possible bias in ratio estimators," below).

**Uncertainty from accidental variation.** Under the conditions stated above, the margin of uncertainty in the estimator  $\bar{x}$  that is attributable to sampling and to small, independent, accidental variations, including random error of measurement (Type III in the next section), may be estimated, for a specified probability, as  $t\hat{\sigma}_{\bar{x}}$ , where  $\hat{\sigma}_{\bar{x}}$  is an



estimator of  $\sigma_x$ . The factor  $t$  depends on the probability level selected for the margin of uncertainty (which will in turn depend on the risks involved) and also on the number of degrees of freedom in the estimator  $\hat{\sigma}_x$ . In large samples the distributions of most estimators are nearly normal, except for frames that exhibit very unusual statistical characteristics. The standard deviation,  $\sigma_x$ , then contains nearly all the information regarding the margin of uncertainty of  $\bar{x}$  that is attributable to accidental variation. Presentation of the results of a survey requires careful consideration when there is reason to question the approximate normality of estimators (Fisher 1956, p. 152; Shewhart 1939, p. 106).

**Random selection.** It is never safe to assume, in statistical work, that the sampling units in a frame are already mixed. A frame comes in layers that are different, owing to geographic origin or to order of production. Even blood, for example, has different properties in different parts of the body.

A random variable is the result of a random operation. A system of selection that depends on the use, in a standard manner, of an acceptable table of random numbers is acceptable as a random selection. Methods of selection that depend on physical mixing, shuffling, drawing numbers out of a hat, throwing dice, are not acceptable as random, because they have no predictable behavior. Neither are schemes that merely remove the choice of sampling units from the judgment of the interviewer. Pseudo-random numbers, generated under competent skill, are well suited to certain types of statistical investigation [see RANDOM NUMBERS].

### Types of uncertainty in statistical data

All data, whether obtained by a complete census or by a sample, are subject to various types of uncertainty. One may reduce uncertainties in data by recognizing their existence and taking steps for improvement in future surveys. Sample design is an attempt to strike an economic balance between the different kinds of uncertainty. There is no point, for example, in reducing sampling error far below the level of other uncertainties.

**Three types of uncertainty.** The following discussion will differentiate three main types of uncertainty.

**Type I.** Uncertainty of Type I comprises built-in deficiencies, or structural limitations, of the frame, questionnaire, or method of test.

Any reply to a question, or any record made by an instrument, is only a response to a stimulus. What stimulus to apply is a matter of judgment. Deficiencies in the questionnaire or in the method of test may therefore arise from incomplete under-

standing of the problem or from unsuitable methods of investigation. Structural limitations are independent of the size or kind of sample. They are built in: a recanvass will not discover them, nor will calculation of standard errors or other statistical calculations detect them.

Some illustrations of uncertainty of Type I are the following:

(a) The frame may omit certain important segments of the universe.

(b) The questionnaire or method of test may fail to elicit certain information that is later found to be needed. The questionnaire may contain inept definitions, questions, and sequences. Detailed accounting will give results different from those given by mere inquiry about total expenditure of a family for some commodity; date of birth gives a different age from that given in answer to the simple question, How old are you? There may be differential effects of interviews depending on such variables as sex and race of the interviewer.

(c) Use of telephone or mail may yield results different from those obtained by personal interview.

(d) Judgments of coders or of experts in the subject matter may differ.

(e) The date of the survey has an important effect on some answers.

**Type II.** Uncertainty of Type II includes operational blemishes and blunders—for example:

(f) One must presume the existence of errors of a noncanceling nature (persistent omission of sampling units designated, persistent inclusion of sampling units not designated, persistent favor in recording results).

(g) One must presume the existence of bias from nonresponse.

(h) Information supplied by coders for missing or illegible entries may favor high or low values.

(i) There may be a large error, such as a unique blunder.

**Type III.** Uncertainty of Type III is caused by random variation. Repeated random samples drawn from the same frame will give different results. Besides, there are inherent uncorrelated, nonpersistent, accidental variations of a canceling nature that arise from inherent variability of investigators, supervisors, editors, coders, punchers, and other workers and from random error of measurement.

**Standard error of an estimator.** The standard error of a result includes the combined effects of all kinds of random variation, including differences within and between investigators, supervisors, coders, etc. By proper design, however, it is possible to get separate estimates of some of these differences.

A small standard error of a result signifies

(i) that the variation between repeated samples will be small and (ii) that the result of the sample agrees well with the equal complete coverage of the same frame. It usually tells little about uncertainties of Type II and never anything about uncertainties of Type I.

**Limitations of statistical inference.** Statistical inference (estimates, standard errors, statistical tests) refers only to the frame that was sampled and investigated. No statistical calculation can, by itself, detect or measure nonsampling errors, although side experiments or surveys may be helpful. No statistical calculation can detect defects in the frame. No statistical calculation can bridge the gap between the frame covered and the universe. This is as true of probability sampling as it is of judgment sampling, and it is true for a complete census of the frame as well.

**Comparison of surveys.** Substantial differences in results may come from what appear to be inconsequential differences in questionnaires or in methods of hiring, training, and supervision of interviewers and coders or in dates of interviewing. The sampling error in a sample is thus not established by comparison against a complete census unless the complete census is the equal complete coverage for the sample.

**Recalls on people not at home.** Many characteristics of people that are not at home at first call, or that are reluctant to respond, may be very different from the average. What is needed is response from everyone selected, including those that are hard to get. To increase the initial size of the sample is no solution. Calculations that cover a wide variety of circumstances show that the amount of information per dollar expended on a survey increases with the number of recalls, the only practicable limit being the time for the completion of the survey. Good sample design therefore specifies that four to six well-timed recalls be made or specifies that recalls continue until the level of response reaches a prescribed proportion. Special procedures, such as intensive subsampling of those not at home on the first or second call, have been proposed (see Leven 1932; Hansen & Hurwitz 1946; Deming 1960).

**Surveys by post.** One can often effect important economies by starting with a mail survey of a fairly large sample properly drawn from a given frame, then finishing with a final determined effort in the form of personal interviews on all or a fraction (one in two or one in three) of the people that failed to reply (Leven 1932). Mail surveys require a frame, in the form of a list of names with reasonably accurate addresses, and provision for keeping records of mailings and of returns.

They are therefore especially adaptable to surveys of members of a professional society, subscribers to a journal, or subscribers to a service. [For further discussion of mail surveys, see ERRORS, article on NONSAMPLING ERRORS.]

### Simple designs for enumerative purposes

The aim in an *enumerative* study is to count the number of people in an area that have certain characteristics or to estimate a quantity, perhaps their annual income, regardless of how they acquired these characteristics. The aim in an *analytic* study is to detect differences between classes or to measure the effects of different treatments.

For illustration consider a study of schizophrenics. One enumerative aim might be to estimate the number of children born to schizophrenic parents before onset of the disease or before the first admission of one of the parents to a hospital for mental diseases. Further aims of the same study might be analytic, such as to discover differences in fertility or in duration of hospitalization caused by different treatments, differences between communities, or differences between time periods.

The finite multiplier typified by  $1/n - 1/N$  (to be seen later) appears in estimators for enumerative purposes. It has no place in estimators for analytic purposes.

Optimum allocation of effort for an enumerative aim may not be optimum for an analytic aim. Moreover, what is optimum for one enumerative characteristic may not be optimum for another. Hence, it will usually be necessary to compromise between competitive aims.

Enumerative aims will occupy most of the remaining space in this article.

The theory presented in this section is for the design commonly called *simple random sampling*. This is often a practicable design, and the theory forms a base for more complex designs.

**A simple procedure of selection and some simple estimators.** Definitions of "frame," "sample," and other terms were introduced above. In addition, it will be convenient to define the *coefficient of variation*. For the  $x$ -population and  $y$ -population of the frame, the coefficients of variation are defined as

$$(12) \quad C_x = \frac{\sigma_x}{\bar{x}}, \quad C_y = \frac{\sigma_y}{\bar{y}}, \quad a > 0, b > 0.$$

In like manner, the symbol  $C_x$  denotes the coefficient of variation of the empirical or theoretical distribution of the random variable  $x$ . The square,  $C_x^2$ , of the coefficient of variation  $C_x$  is called the *rel-variance* of  $x$ . The coefficient of variation is especially useful for characteristics (such as height)



that are positive. It is often helpful to remember, for example, that  $C_x = C_y = C_z$  because  $x$ ,  $\bar{x}$ , and  $X$  are constant multiples of each other.

The procedure of selection specified earlier gives every member of the frame the same probability of selection as every other member, wherefore

$$(13) \quad \begin{aligned} E\bar{x} &= a, \\ E\bar{y} &= b. \end{aligned}$$

That is,  $\bar{x}$  and  $\bar{y}$  are unbiased estimators of  $a$  and  $b$ , respectively. Moreover,

$$(14) \quad \begin{aligned} X &= N\bar{x}, \\ Y &= N\bar{y} \end{aligned}$$

are unbiased estimators of  $A$  and  $B$ .

Often, a ratio such as

$$(15) \quad \phi = \frac{A}{B} = \frac{a}{b}$$

is of special interest. The sample gives

$$(16) \quad f = \frac{X}{Y} = \frac{x}{y} = \frac{\bar{x}}{\bar{y}}$$

as an estimator of  $\phi$ . If the total  $y$ -population,  $B$ , is known from another source, such as a census,  $A$  may be estimated by the formula

$$(17) \quad X' = Bf.$$

This estimator  $X'$  is called a *ratio estimator*. It will be more precise than the estimator  $X = N\bar{x}$  in (14) if the correlation between  $x_i$  and  $y_i$  is high. Other estimators will be discussed later (for example, regression estimators). Theory provides a basis for the choice of estimator.

**Possible bias in ratio estimators.** Necessary and sufficient conditions for there to be no bias in  $f$  as an estimator of  $\phi$  are that  $Ey \neq 0$  and that  $x_i/y_i$  and  $y_i$  be uncorrelated—that is, that  $E[(x/y)y] = E(x/y)Ey$ . In practice, if bias exists at all, it is usually negligible when the sample contains more than three or four sampling units.

**Sampling with and without replacement.** Usually, in the sampling of finite populations, one permits a sampling unit to come into the sample only once. In statistical language, this is sampling *without replacement*. Tests of physical materials are sometimes destructive, and a second test would be impossible. To draw without replacement, one simply disregards a random number that appears a second time (or uses tables of so-called random permutations).

There are circumstances, however, in which one accepts the random numbers as they come and permits a sampling unit to come into the sample more than once. This is sampling *with replacement*.

Hereafter, most equations will be written for

sampling without replacement. It is a simple matter to drop the fraction  $1/N$  from any formula to get the corresponding formula for sampling with replacement. Actually, in practice, samples are usually such a small part of the frame that the fraction  $1/N$  is ignored, even though the sampling be done without replacement.

**Variances.** The variances of the estimator  $\bar{x}$  derived from the sampling procedure described earlier are

$$(18) \quad \left\{ \begin{array}{ll} \text{without} & \sigma_{\bar{x}}^2 = \frac{N-n}{N-1} \frac{\sigma^2}{n} \\ \text{replacement} & \approx \left( \frac{1}{n} - \frac{1}{N} \right) \sigma^2, \\ \\ \text{with} & \sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}. \\ \text{replacement} & \end{array} \right.$$

(The sign  $\approx$  indicates an approximation that is sufficiently close in most practice.)

Similar expressions hold for  $\bar{y}$ . For the ratio  $f = \bar{x}/\bar{y}$ , the approximation

$$(19) \quad C_f^2 \approx \left( \frac{1}{n} - \frac{1}{N} \right) (C_x^2 + C_y^2 - 2abC_{xy})$$

is useful if  $n$  be not too small. Here

$$(20) \quad C_{xy} = \frac{1}{Nab} \sum (a_i - a)(b_i - b)$$

is the rel-covariance of the  $x$ -population and  $y$ -population per sampling unit in the frame.

When the ratio estimator of the total  $x$ -population is derived as in eq. (17), eq. (19) gives the same approximation for  $C_{f'}^2$ .

**Estimate of aggregate characteristic—number of units in class unknown.** It often happens in practice that one wishes to estimate the aggregate value of some characteristic of a subclass of a group when the total number of units in the subclass is unknown. For example, one might wish to estimate the aggregate income of women aged 15 or over that live in a certain district, are gainfully employed, and have at least one child under 12 years old at home (this specification defines the universe). The number of women that meet this specification is not known. An estimate of the average income per woman of this specification, prepared from a sample, suffers very little from this gap in available knowledge, but an estimate of the total income of all such women is not so fortunate.

As an illustration, suppose that the frame is a serialized list of  $N$  women aged 15 or over and that the sample is a simple random sample of  $n$  of these women, drawn with replacement by reading out  $n$  random numbers between 1 and  $N$ . Informa-

tion on the  $n$  women is collected, and it is noted which ones belong to the specified subclass—that is, which ones live in a certain district, are gainfully employed, and have at least one child under 12 years old at home. Suppose that this number is  $n_s$  and that the average income of the  $n_s$  women is  $\bar{x}_s$ . Of course,  $n_s$  is a random variable with a binomial distribution.

What is the rel-variance of  $\bar{x}_s$ ? Let  $C_s^2$  be the rel-variance between incomes of the women in the frame that belong to the subclass. It is a fact that the conditional rel-variance of  $\bar{x}_s$ , for samples of size  $n_s$  of the specified subclass, will be  $C_s^2/n_s$ , just as if the women of this subclass had been set off beforehand into a separate stratum and a sample of size  $n_s$  had been drawn from it.

The conditional expected value of  $\bar{x}_s$  over all samples of fixed size  $n_s$  in the subclass has moreover the convenient property of being the average income of all the women in the frame that belong to this subclass. It is for this reason that the conditional rel-variance of  $\bar{x}_s$  is useful for assessing the precision of a sample at hand. For purposes of design, one uses the rel-variance of  $\bar{x}_s$  over all samples of size  $n$ , which is  $C_s^2 E(1/n_s)$ , or very nearly  $C_s^2[1 + Q/nP]/nP$ , where  $P$  is the proportion of all women 15 or over that meet the specification of the subclass, and  $P + Q = 1$ .

In contrast, any estimator,  $X_s$ , of the aggregate income of all the women in the specified subclass will not have such convenient properties as  $\bar{x}_s$ . The conditional expectation of  $X_s$ , for samples of size  $n_s$ , is not equal to the aggregate income of all the women in the frame that belong to the subclass. The conditional rel-variance of  $X_s$  for a sample of size  $n_s$  at hand, although equal to the conditional rel-variance of  $\bar{x}_s$ , therefore requires careful interpretation. Instead of attempting to interpret the conditional rel-variance of  $X_s$ , one may elect to deal with the variance of  $X_s$  in all possible samples of size  $n$ . Thus, if  $X_s$  is set equal to  $(N/n)n_s\bar{x}_s$  (here  $N/n$  is used as an expansion factor equal to the reciprocal of the probability of selection), it is a fact that the rel-variance of  $X_s$  over all samples of size  $n$  will be  $(C_s^2 + Q)/nP$  (see Deming 1960, p. 129).

The problem with  $X_s$  arises from the assumption that  $N_s$ , the number of women in the frame that meet the specification of the subclass, is unknown. If  $N_s$  were known, one could form the estimator  $X_s = N_s\bar{x}_s$ , which would have all the desirable properties of  $\bar{x}_s$ .

One way to reduce the variance of the total income,  $X_s$ , of the specified class is (1) to select from the frame a large preliminary sample, (2) by

an inexpensive investigation to classify the units of the preliminary sample into two classes, those that belong to the specified class and those that do not, (3) to investigate a sample of the units that fell into the specified class, to acquire information on income. The preliminary sample provides an estimate of  $N_s$ , and the final sample provides an estimate of  $\bar{x}_s$ . The product gives the estimate  $X_s = N_s\bar{x}_s$  for the total income in the specified class. (For the variance of  $X_s$  and for optimum sizes of samples, see Hansen, Hurwitz, & Madow, 1953, vol. 1, pp. 65 and 259.)

If, further,  $N$  were not known and only the probability,  $\pi$ , of selection, to be applied to every sampling unit in the frame, were known, both  $n$  and  $n_s$  will be random variables, and there will be a further inflation of the rel-variance of any estimator of the aggregate income of all the women in the specified subclass. Thus, if  $X_s$  be set equal to  $n_s\bar{x}_s/\pi$  for such an estimator, then the unconditional rel-variance of  $X_s$  will be  $(C_s^2 + 1)/nP$ . The conditional rel-variance of  $\bar{x}_s$ , however, is still  $C_s^2/n_s$ .

It may be noted that for a small subclass there is little difference between  $C_s^2 + Q$  and  $C_s^2 + 1$ .

Examples are common. Thus, one might read out a two-digit random number for each line of a register, following the rule that the item listed on a line will be drawn into the sample if the random number is 01. If counts from outside sources are not at hand or are not used, then the rel-variance of an estimator,  $X_s$ , of the total number or total value of any subclass of items on the register contains the factor  $C_s^2 + 1$ .

*Use of thinning digits.* Reduction of the probability of selection of units of specified characteristics (such as items of low value) through the use of thinning digits may produce either the factor  $C_s^2 + Q$  or the factor  $C_s^2 + 1$  in the rel-variance of an estimator of an aggregate, depending on the mode of selecting the units.

*Estimates of variances.* Estimates of variances are supplied by the sample itself, under proper conditions, as was discussed above. Some of the more important estimators follow, denoted by a circumflex ( $\hat{\phantom{x}}$ ). For the variance of  $\bar{x}$ ,

$$(21) \quad \hat{\sigma}_{\bar{x}}^2 = \left(\frac{1}{n} - \frac{1}{N}\right) \frac{1}{n-1} \sum (x_i - \bar{x})^2,$$

with a similar expression for  $\hat{\sigma}_{\bar{y}}^2$ . For the covariance,

$$(22) \quad \hat{\sigma}_{\bar{x}\bar{y}} = \left(\frac{1}{n} - \frac{1}{N}\right) \frac{1}{n-1} \sum (x_i - \bar{x})(y_i - \bar{y}).$$

Eqs. (21) and (22), with  $N$  infinite, were developed by Gauss (1823). These estimators are un-



biased;  $\sqrt{\hat{\sigma}_x^2}$  is a slightly biased estimator of  $\sigma_x$ , but the bias is negligible for  $n$  moderate or large.

Under almost all conditions met in practice, one may set

$$(23) \quad t = \frac{\bar{x} - E\bar{x}}{\hat{\sigma}_x}$$

and compare this quantity with tabulated values of  $t$  to find the margin of uncertainty in  $\bar{x}$  for any specified probability. Such calculations give excellent approximations unless the distribution of sampling units in the frame is highly skewed. Extreme skewness may often be avoided by stratification (discussed below).

A useful approximate estimator for the rel-variance of  $f = X/Y = \bar{x}/\bar{y}$  is

$$(24) \quad \hat{C}_f^2 \approx \left(\frac{1}{n} - \frac{1}{N}\right) \frac{1}{(\bar{n} - 1)\bar{x}^2} \sum (x_i - f y_i)^2.$$

This formula is derived by combination of eqs. (19), (21), and (22). In accordance with a previous remark, one may take  $\hat{C}_{X'} = \hat{C}_f$ , where  $X'$  is the ratio estimator of  $A$  as given by (17).

**Size of frame usually not important.** Because of the way in which  $N$  enters the variances, the size of the frame has little influence on the size of sample required for a prescribed precision, unless the sample is 20 per cent or more of the total frame. For instance, the sample required to reach a specified precision would be the same for the continental United States as for the Boston Metropolitan Area, on the assumption that the underlying variances encountered are about the same for the entire United States as for Boston.

**Special form for attributes (0,1 variate).** In many studies a sampling unit gives only one of two possible observations, such as *yes* or *no*, *male* or *female*, *heads* or *tails*. The above equations then assume a simple form.

If each person in a frame is a sampling unit, and if  $a_i = 1$  for *yes*,  $a_i = 0$  for *no*, then the total  $x$ -population,  $A$ , in the frame is the total number of *yes* observations that would be recorded in the equal complete coverage, and  $a$  is the proportion *yes*, commonly denoted by  $p$ . The variance between the  $a_i$  in the frame is

$$(25) \quad \sigma_a^2 = pq,$$

where  $p + q = 1$ .

The random variate,  $x_i$ , will take the value 0 or 1;

$$(26) \quad x = \sum x_i$$

will be the number of *yes* observations in the sample, and

$$(27) \quad \hat{p} = x/n$$

will be the proportion *yes* in the sample. Replacement of  $\bar{x}$  by  $\hat{p}$  in previous equations shows that  $\hat{p}$  is an unbiased estimator of the proportion *yes* in the frame and that

$$(28) \quad \sigma_{\hat{p}}^2 = \left(\frac{1}{n} - \frac{1}{N}\right) pq.$$

It is important to note that this variance is valid only if each sampling unit produces the value 0 or 1. It is not valid, for instance, for a sample of segments of area if there is more than one person per segment, or if the segments are clustered (as discussed below).

For an estimate of the variance of  $\hat{p}$  (provided the sampling procedure meets the conditions stated) one may use

$$(29) \quad \hat{\sigma}_{\hat{p}}^2 = \left(\frac{1}{n} - \frac{1}{N}\right) \hat{p}\hat{q},$$

where  $\hat{p} + \hat{q} = 1$ .

**How good is an estimator of a variance?** The variance of the estimator  $\hat{\sigma}_x^2$  in eq. (21) depends on the standardized fourth moment,  $\beta_2$ , of the frame and on the number of degrees of freedom for the estimator. Thus, if one defines

$$(30) \quad \beta_2 = \frac{1}{N} \sum \left( \frac{a_i - a}{\sigma} \right)^4,$$

then the rel-variance of the estimator  $\hat{\sigma}_x^2$  of eq. (21) will be  $(\beta_2 - 1)/n$ , which diminishes with  $n$ .

**Systematic selection.** A simple and popular way to spread the sample over the frame is to select every  $Z$ th unit, with a random start between 1 and  $Z$ , where  $Z = N/n$ . This is called systematic sampling with a single random start, and it is one form of patterned sampling. In certain kinds of materials, specifically those in which nearby sampling units are, on the average, more similar than units separated by a longer interval, systematic sampling will be slightly more efficient than stratified random sampling (Cochran 1946).

A disadvantage of systematic sampling with a single random start is that there is no strictly valid way to make a statistical estimate of the standard error of a result so obtained. This is because the single start is equivalent to the selection of only one sampling unit from the  $Z$  possible sampling units that could be formed. One may nevertheless, under proper conditions, get a useful approximation to the rel-variance by using the sum of squares of successive pairs. Eq. (21) with  $n = 2$  and  $N = Z$  gives the estimator

$$(31) \quad \hat{C}_x^2 = \left(1 - \frac{2}{Z}\right) \frac{\sum (x_{i1} - x_{i2})^2}{[\sum (x_{i1} + x_{i2})]^2},$$

where the summation runs over all pairs.

Hidden and unsuspected periodicities often turn up, and in such cases the above formula may give a severe underestimate or overestimate of the variance. For example, every  $n$ th household might be nearly in phase with the natural periodicity of income, rent, size of family, and other characteristics associated with corners and with the configuration of dwelling units within areas and within apartment houses. Systematic sampling of physical elements or of time intervals can lead to disaster.

A statistician will therefore justify a single random start and use of eq. (31) only if he has had long experience with a body of material.

Instead of a single random start between 1 and  $N/n$ , one may take two random starts between 1 and  $2N/n$  and every  $(2N/n)$ th sampling unit thereafter. Extension to multiple random starts is obvious. Two or more random starts give a valid estimate of the variance. Fresh random starts every six or eight zones will usually reap any possible advantage of systematic sampling and will avoid uncertainty in estimation of the variance.

**Efficiency of design.** The relative efficiency of two sampling procedures, I and II, that give normally distributed estimators of some characteristic is by definition the ratio of the inverses of the variances of these estimators for the same size,  $n$ , of sample. In symbols ( $E$  denotes efficiency),

$$(32) \quad \frac{E_I}{E_{II}} = \frac{\sigma_{II}^2}{\sigma_I^2}, \quad n_I = n_{II}.$$

This concept of efficiency is due to Fisher (1922).

Comparison of costs is usually more important than comparison of numbers of cases. Let the costs be  $c_I$  and  $c_{II}$  for equal variances. Then

$$(33) \quad \frac{E_I}{E_{II}} = \frac{c_{II}}{c_I}, \quad \sigma_I = \sigma_{II}.$$

Comparison of efficiencies of estimators whose distributions depart appreciably from normality require special consideration.

**Sampling moving populations.** A possible procedure in sampling moving populations is to count and tag all the people visible from a number of enumerators' posts through a period of a day or a week (the first round) and then to repeat the count from the same or different posts some time later (the second round). The  $n_1$  people counted and tagged in the first round constitute a mobile frame for the second round. If the number of people counted in the first round is  $n_1$ , and if the number counted in the second round is  $n_2$ , with an intersection of  $n_{12}$  for people counted in both rounds, then an estimator of the total number of mobile inhabitants in the whole area is  $\hat{N} = n_1 n_2 / n_{12}$  (Yates 1949, p. 43; Deming & Keyfitz 1967).

## More complex designs

**Considerations of cost—clustering.** The total cost of a survey includes cost of preparing the frames and cost of travel to the units selected. In some surveys it may be possible to get more information per unit cost by enlarging the sampling unit, a procedure commonly called *clustering*. One may, for example, define a sampling unit as comprising all the dwellings in a compact segment of area. Further, one may, with experience and care, subsample dwelling units from a selected cluster or select one member of a family where two or more members qualify for the universe. Again, in a national survey, one may restrict the sample to a certain number of counties that will come into the sample by a random process. Or, in a survey of a city, one may restrict the sample to a random selection of blocks.

Any such plan reduces the interviewer's expenses for travel and reduces the cost of preparing the frame. However, restriction of the sample usually also increases variances, unless the total number of households in the sample be increased as compensation. It should be remembered, though, that the actual precision obtained by the use of cluster sampling may be nearly as good as that obtained by an unrestricted random selection of the same number of dwelling units with no clustering.

Theory indicates the optimum balance between enlargement of the sampling unit and the number of sampling units to include in the sample. Obviously, the theory is more complex than that discussed in the last section. Stratification, ratio estimators, and regression estimators are additional techniques that, under certain conditions, yield further increases in efficiency (see below).

**An example.** The following illustration refers to a sample of a city: (i) Suppose that it has been determined in advance that for the main purposes of a survey the optimum size of areal unit is a compact group of five dwelling units, called a segment. (ii) A sampling unit within the city will consist of  $\bar{n}$  segments from a larger number of segments contained in a block. The  $\bar{n}$  segments of a sampling unit (if  $\bar{n} > 1$ ) should be scattered over the block. A good way to effect this scatter is by a systematic selection. (iii) The  $m$  sampling units in the city will be selected by random numbers. For simplicity, assume that all blocks in the city contain an equal number,  $\bar{N}$ , of segments. Suppose that there are

$M$  blocks in the whole city,

$\bar{N}$  segments in a block,

$N = M\bar{N}$  segments in the whole city,



$\bar{n}$  segments in a sampling unit,  
 $\bar{N}/\bar{n}$  sampling units in a block,  
 $M\bar{N}/\bar{n}$  or  $N/\bar{n}$  sampling units in the whole city,  
 $m$  sampling units in the sample.

Then if

$$(34) \quad \bar{x} = x/m\bar{n},$$

one may take

$$(35) \quad X = N\bar{x}$$

for an estimator of the  $x$ -population in the whole city. For this estimator,

$$(36) \quad C_x = C_{\bar{x}},$$

and

$$(37) \quad \text{var } \bar{x} = \frac{N - m\bar{n}}{N - \bar{n}} \left( \frac{\sigma_b^2}{m} + \frac{\bar{N} - \bar{n}}{\bar{N} - 1} \frac{\sigma_w^2}{m\bar{n}} \right).$$

If  $m$  is small compared with  $M$ ,

$$(38) \quad \text{var } \bar{x} \cong \frac{\sigma_b^2}{m} + \frac{\bar{N} - \bar{n}}{\bar{N} - 1} \frac{\sigma_w^2}{m\bar{n}}.$$

If, also,  $\bar{n}$  is small compared with  $\bar{N}$ ,

$$(39) \quad \text{var } \bar{x} \cong \frac{\sigma_b^2}{m} + \frac{\sigma_w^2}{m\bar{n}}.$$

Here  $\sigma_b^2$  is the variance between blocks of the mean  $x$ -population per sampling unit, and  $\sigma_w^2$  is the average variance between sampling units within blocks.

**Important principle in size of secondary unit.** Suppose that the cost of adding one more block to the sample is  $c_1$  (cost of maps, preparation, delineation of segments, travel) and that the cost of an interview in an additional sampling unit is  $c_2$ . Then the total cost of the survey will be

$$(40) \quad K = mc_1 + m\bar{n}c_2.$$

In eq. (37)  $\text{var } \bar{x}$  will be at its minimum for a fixed cost  $K$  if

$$(41) \quad \bar{n} = \frac{\sigma_w}{\sigma_b} \sqrt{\frac{c_1}{c_2}}, \quad \text{optimum } \bar{n}.$$

This equation was derived by both L. H. C. Tippett and Shewhart, independently, in 1931.

Note that  $m$  does not appear in this equation. That is, the optimum value of  $\bar{n}$  on the basis of the cost function (40) is independent of  $m$ , the number of sampling units in the sample (and very nearly independent of the number of blocks in the sample).

The optimum  $m$  is found by substituting the optimum  $\bar{n}$  from eq. (41) into eq. (40) and solving for  $m$ . Of course, it is necessary to assume values for  $\sigma_w/\sigma_b$  and for  $\sqrt{c_1/c_2}$  to do this. (Be-

cause each sampling unit will usually fall in a different block,  $m$  will usually be exactly or nearly as large as the number of blocks in the sample.) Usual numerical values of  $\sigma_w/\sigma_b$  and of  $c_1/c_2$  lead to small values of  $\bar{n}$  and to large values of  $m$ . Efficient design therefore usually requires a small sample from a block and dispersion of the sample into a large number of blocks.

Extension of this theory to a national sample, and to stratified designs and ratio estimators, leads to the same principle.

Variation in size of segment will increase  $\text{var } \bar{x}$  by the factor  $1 + C_n^2/\bar{n}$ , where  $C_n^2$  is the rel-variance of the distribution of the number of dwelling units per segment. A similar factor,  $1 + C_{N_i}^2/m$ , measures the increase in  $\text{var } \bar{x}$  from variation in the number of segments per block.

**Replicated designs for ease in estimation of variance.** Replication of a sample in two or more interpenetrating networks of samples will provide a basis for rapid calculation of a valid estimate of the standard error of any result, regardless of the complexity of the procedure of selection and of the formulas for the formation of estimates [Mahalanobis 1944; Deming 1950; 1960; see also INDEX NUMBERS, article on SAMPLING].

### Stratified sampling

The primary aim of stratified sampling is to increase the amount of information per unit of cost. A further aim may be to obtain adequate information about certain strata of special interest.

One way to carry out stratification is to rearrange the sampling units in the frame so as to separate them into classes, or strata, and then to draw sampling units from each class. The goal should be to make each stratum as homogeneous as possible, within limitations of time and cost. Stratification is equivalent to blocking in the design of an experiment. It is often a good plan (i) to draw a preliminary sample from the frame without stratification; (ii) to classify into strata the units in the preliminary sample, and (iii) to draw, for the final sample, a prescribed number of sampling units from each stratum so formed. Step (i) will sometimes require an inexpensive investigation or test of every sampling unit in the preliminary sample to determine which stratum it belongs to.

Stratification is one way to make use of existing information concerning the frame other than the information obtained from investigating the sampling units in the final sample itself. Other ways to use existing information are through ratio estimators and regression estimators (see below).

In practice a frame is to some extent naturally

Table 2 — Notation and definitions for the frame ( $M = 2$  strata)

STRATUM	NUMBER OF SAMPLING UNITS		STRATUM'S PROPORTION OF SAMPLING UNITS IN THE FRAME	POPULATION		BETWEEN THE POPULATIONS OF THE SAMPLING UNITS WITHIN THE STRATUM	
	In the frame	In the sample		Average per sampling unit in the stratum	Total in the stratum	Standard deviation	Variance
1	$N_1$	$n_1$	$P_1 = \frac{N_1}{N}$	$a_1$	$A_1 = N_1 a_1$	$\sigma_1$	$\sigma_1^2$
2	$N_2$	$n_2$	$P_2 = \frac{N_2}{N}$	$a_2$	$A_2 = N_2 a_2$	$\sigma_2$	$\sigma_2^2$
Total for the frame	$N$	$n$	1	—	$A$	—	—
Unweighted average per stratum	$\bar{N} = \frac{N}{M}$	$\bar{n} = \frac{n}{M}$	$\frac{1}{M}$	—	$\bar{A} = \frac{A}{M}$	—	—
Weighted average per sampling unit	—	—	—	$a = \frac{A}{N}$	—	$\sigma_w$	$\sigma_w^2$

Source: Deming 1960, p. 286.

stratified to begin with. Thus, areas in geographic order usually are already pretty well stratified in respect to income, occupation, density of population, tastes of the consumer, and other characteristics. No frame arrives thoroughly mixed, and any plan of sampling should be applied by zones, so as to capture the natural stratification. Theory serves as a guide to determine whether further stratification would be profitable.

Plans of stratification for enumerative studies. Several plans of stratified sampling for enumerative studies will now be described.

The notation and definitions to be used in this discussion are given in tables 2 and 3. (Note that  $\bar{N}$  and  $\bar{n}$  are defined differently here than they were earlier.) These tables are presented in terms of two strata ( $M = 2$ ), but extension to a greater number of strata follows obviously. The following additional definitions are needed:

$$(42) \quad \sigma_R^2 = Q_1 \sigma_1^2 + Q_2 \sigma_2^2 + Q_3 \sigma_3^2,$$

the average reverse variance between sampling units within strata, and

$$(43) \quad \bar{\sigma}_R = Q_1 \sigma_1 + Q_2 \sigma_2 + Q_3 \sigma_3,$$

the average reverse standard deviation between sampling units within strata, where  $Q_1 + P_1 = 1$ .

*Plan A (no stratification):* The scheme of sampling described above will be designated plan A. It is needed here for comparison, and also because it constitutes the basis for selection from any stratum.

Note that in plan A, as in plans B, D, F, and H, below, all the sampling units in the frame have equal probability of selection, namely  $n/N$ , wherefore  $E\bar{x} = a$  and  $EX = A$ .

*P<sub>1</sub> known—whole frame classified.* Two sampling plans for which the proportions in each stratum are known (or ascertainable) and the the whole frame is classified will now be described.

*Plan B (proportionate sampling):* Decide with the help of eq. (47) the size,  $n$ , of the sample required. Compute next

$$(44) \quad n_i = nN_i/N = nP_i.$$

Draw by random numbers, as in plan A, a sample of size  $n_i$  from stratum  $i$ . Investigate every member of the sample, and calculate

$$(45) \quad X_1 = N_1 x_1/n_1, \quad X_2 = N_2 x_2/n_2, \quad \text{etc.}$$

Table 3 — Notation and definitions for the sample

Stratum	Population in the sample	Mean population per sampling unit	Estimated total population	Variance of this estimator*
1	$x_1$ x-population in stratum 1	$\bar{x}_1 = \frac{x_1}{n_1}$	$X_1 = N_1 \frac{x_1}{n_1}$	$\text{var } X_1$
2	$x_2$ x-population in stratum 2	$\bar{x}_2 = \frac{x_2}{n_2}$	$X_2 = N_2 \frac{x_2}{n_2}$	$\text{var } X_2$
Sum	$x$	—	$X$	$\text{var } X$

\* The variances are additive only if the  $N_i$  (or  $P_i$ ) are known and used in the estimator  $X$ .



(For simplicity, most formulas will henceforth be written for two strata, in conformance with tables 2 and 3. Extension to more strata is obvious.) Here,  $n_i/N_i = n/N$ , wherefore

$$\begin{aligned} X &= X_1 + X_2 \\ &= \frac{N}{n} (x_1 + x_2) \\ &= N \frac{\bar{x}}{n} = N\bar{x} \end{aligned} \quad (46)$$

and

$$\text{var } \bar{x} = \left( \frac{1}{n} - \frac{1}{N} \right) \sigma_w^2, \quad \text{plan B.} \quad (47)$$

The  $n_i$  of eq. (44) and later expressions will not in general be integers. In practice one uses the closest integer; the effects on variance formulas are usually completely negligible.

*Plan C (Neyman sampling):* Decide with the help of eq. (49) the size,  $n$ , of the sample required. Compute next the Neyman allocation (Neyman 1934),

$$n_i = n P_i \sigma_i / \bar{\sigma}_w. \quad (48)$$

Draw by random numbers, as in plan A, a sample of size  $n$ , from stratum  $i$ . Investigate every member of the sample. Form estimators  $X_1$ ,  $X_2$ , and  $X = X_1 + X_2$ . Form  $\bar{x} = X/N$  for an unbiased estimator of  $a$ . Here

$$\text{var } \bar{x} = \frac{(\bar{\sigma}_w)^2}{n} - \frac{\sigma_w^2}{N}, \quad \text{plan C.} \quad (49)$$

The Neyman allocations are the optimal  $n_i$  for minimizing  $\text{var } \bar{x}$  when the  $P_i$  are known.

*$P_i$  known—only a sample classified.* One may, in appropriate circumstances, require only the classification of a preliminary sample drawn from the frame. The decision hinges on the costs of classification and the expected variances of the plans under consideration.

*Plan D:* Decide with the help of eq. (50) the size,  $n$ , of the sample required. Draw the sample as in plan A. Classify the sampling units into strata. The number,  $n_i$ , of sampling units drawn from stratum  $i$  will be a random variable. Carry out the investigation of every unit of the sample. Form  $X_1$ ,  $X_2$ ,  $X$ , and  $\bar{x}$  as in plan B. Then

$$\text{var } \bar{x} = \left( \frac{1}{n} - \frac{1}{N} \right) \left( \sigma_w^2 + \frac{1}{n} \sigma_n^2 \right), \quad \text{plan D.} \quad (50)$$

*Plan E:* Decide with the help of eq. (52) the size,  $n$ , of the final sample. Draw by random numbers a preliminary sample of size  $n'$ . Thin (reduce)

by random numbers the strata of the preliminary sample to reach the Neyman ratios

$$(51) \quad \frac{n_1}{n'_1} : \frac{n_2}{n'_2} : \dots = \sigma_1 : \sigma_2 : \dots$$

and simultaneously the total sample,  $n$ . Here  $n'_1$ ,  $n'_2$ , etc., are the sizes of the preliminary sample in the several strata, and  $n_1$ ,  $n_2$ , etc., are the sizes of the final sample. For greatest economy, choose  $n'$  so that one stratum will require no thinning. Carry out the investigation of every unit of the final sample. Form the estimators  $X_1$ ,  $X_2$ , and  $X = X_1 + X_2$ . Then  $\bar{x} = X/N$  will again be an unbiased estimator of  $a$ , but now

$$\begin{aligned} \text{var } \bar{x} &= \overbrace{\frac{(\bar{\sigma}_w)^2}{n} - \frac{\sigma_w^2}{N}}^{\text{plan C}} + \frac{1}{n} \left( \frac{1}{n'} - \frac{1}{N} \right) \bar{\sigma}_w \bar{\sigma}_n \\ &\approx \frac{1}{n} \left[ (\bar{\sigma}_w)^2 + \frac{1}{n'} \bar{\sigma}_w \bar{\sigma}_n \right], \quad \text{plan E} \end{aligned} \quad (52)$$

the latter form useful if  $N$  is large relative to  $n'$ .

*Sequential classification of units into strata.* We now describe two plans in which the sample-sizes,  $n_i$ , are reached sequentially, with considerable saving under appropriate conditions.

*Plan F:* Determine the desired sample-sizes,  $n_i$ , as in plan B. Draw by random numbers one unit at a time from the frame, and classify it into its proper stratum. Continue until the quotas,  $n_i$ , are all filled. Form  $X$  as in plan B;  $\text{var } \bar{x}$  will be the same as for plan B.

*Plan G:* This is the same as plan F except that the sample sizes,  $n_i$ , are fixed as in plan C. Form  $X$  as in plan C;  $\text{var } \bar{x}$  will be the same as for plan C.

*$P_i$  not known in advance.* When the proportions,  $P_i$ , in the frame are unknown, estimates thereof must come from a sample, usually a preliminary sample of size  $N' > n$ , where  $n$  is the size of the final sample.

*Plan H:* Decide with the help of eq. (55) the size,  $n$ , for the final sample. Compute the optimum size,  $N'$ , of the preliminary sample by the formula

$$(53) \quad \frac{n}{N'} = \frac{\sigma_w}{\sigma_b} \sqrt{\frac{c_1}{c_2}},$$

where  $c_1$  is the average cost of classifying a sampling unit in the preliminary sample, and  $c_2$  is the average cost of the final investigation of one sampling unit.

The procedure is to draw as in plan A a preliminary sample of size  $N'$  and to classify it into strata. Treat the preliminary sample as a frame of size  $N'$ . Then thin all strata of the preliminary sample proportionately to reach the final total size,  $n$ .

Carry out the investigation of every sampling unit in the final sample. An unbiased estimator of  $\alpha$  is

$$(54) \quad \bar{x} = \frac{x}{n},$$

where  $x$  is the total  $x$ -population in the sample. Then

$$(55) \quad \text{var } \bar{x} \approx \frac{\sigma_w^2}{n} + \frac{\sigma_b^2}{N'} = \frac{\sigma^2}{N'} + \left( \frac{1}{n} - \frac{1}{N'} \right) \sigma_w^2,$$

plan H,

is an excellent approximation if  $N$  be large relative to  $N'$ .

*Plan I:* Decide with the help of eq. (59) the size,  $n$ , for the final sample. Compute the optimum size,  $N'$ , of the preliminary sample, using the equation (Neyman 1938)

$$(56) \quad \frac{n}{N'} = \frac{\bar{\sigma}_w}{\sigma_b} \sqrt{\frac{c_1}{c_2}}.$$

Draw as in plan A a preliminary sample of size  $N'$ . Classify it as in plan H. Thin the strata differentially to satisfy the Neyman ratios

$$(57) \quad \frac{n_1}{N'_1} : \frac{n_2}{N'_2} : \dots = \sigma_1 : \sigma_2 : \dots$$

and to reach the desired final total sample-size,  $n$ . Carry out the investigation of every sampling unit in the final sample. An unbiased estimator of  $\alpha$  is

$$(58) \quad \bar{x} = \frac{1}{N'} \left( \frac{N'_1}{n_1} x_1 + \frac{N'_2}{n_2} x_2 \right),$$

for which

$$(59) \quad \text{var } \bar{x} \approx \frac{(\bar{\sigma}_w)^2}{n} + \frac{\sigma_b^2}{N'}, \quad \text{plan I,}$$

is an excellent approximation if  $N$  be large relative to  $N'$  and to  $n$ .

One may use plan F or plan G in combination with plan H or plan I to reap the benefit of many strata without actually classifying the entire preliminary sample,  $N'$  (Koller 1960).

*Gains of stratified sampling.* Gains of stratified sampling can be evaluated by comparing variances. Denote by  $A$ ,  $B$ , and  $C$  the variances of the estimators of  $\alpha$  calculated by the plans A, B, and C. Then

$$(60) \quad \frac{A - B}{A} = \frac{\sigma^2 - \sigma_w^2}{\sigma^2} = \frac{\sigma_b^2}{\sigma^2} = \sum_{i < j} P_i P_j \left( \frac{a_j - a_i}{\sigma} \right)^2,$$

$$(61) \quad \frac{B - C}{B} \approx \frac{\sigma_w^2 - (\bar{\sigma}_w)^2}{\sigma_w^2} = \sum_{i < j} P_i P_j \left( \frac{\sigma_j - \sigma_i}{\sigma_w} \right)^2.$$

For example, if  $P_1 = .6$ ,  $P_2 = .4$ , and  $\sigma_w^2 = .8\sigma^2$ ,  $(A - B)/A$  would be  $(1 - .8)/1 = .2$ , meaning that 100 interviews selected according to plan B would

give rise to the same variance as 125 selected according to plan A.

The gains of plans F and G over plan A are the same as the gains of plans B and C over plan A. The average gains in repeated trials of plans D and E are less. If  $\sigma_R^2$  and  $\bar{\sigma}_w \bar{\sigma}_R$  are large, plans D and E will usually not be good choices. For large samples, however, in circumstances where  $\sigma_R^2$  and  $\bar{\sigma}_w \bar{\sigma}_R$  are not large, the gains of plans D and E may be almost equal to the gains of plans B and C, at considerably less cost.

Eqs. (60) and (61) show that the gain to be expected from the proposed formation of a new stratum,  $i$ , will not be impressive unless its proportion,  $P_i$ , be appreciable, or unless its  $\sigma_i$  or its  $a_i$  be widely divergent from the average.

*Stratification to estimate over-all ratio.* The case to be used for illustrating stratified sampling to estimate an over-all ratio consists of three strata: stratum 1 for large units (for example, high incomes or large farms), stratum 2 for medium-sized units, and stratum 3 for small units. Here stratum 1 is to be covered 100 per cent; obvious modifications take care of the case in which stratum 1 is not sampled completely.

First take as an estimator of  $\phi$

$$(62) \quad f = \frac{X}{Y} = \frac{A_1 + N_2 \bar{x}_2 + N_3 \bar{x}_3}{B_1 + N_2 \bar{y}_2 + N_3 \bar{y}_3},$$

in the notation of tables 1 and 2, with  $B_i$  as the value of the  $y$ -characteristic in stratum  $i$  of the frame. Optimum allocation to strata 2 and 3 is very nearly reached if both

$$(63a) \quad n_2 = \frac{B_2 s_2}{B_2 s_2 + B_3 s_3}$$

and

$$(63b) \quad n_3 = \frac{B_3 s_3}{B_2 s_2 + B_3 s_3},$$

wherein  $s_2$  and  $s_3$  are the standard deviations of the ratio of  $x$  to  $y$  in strata 2 and 3.

If, as is often the case,  $s_2$  and  $s_3$  do not differ much, or if little is known about them in advance, one can still make an important gain in efficiency by setting  $n_2 : n_3 = B_2 : B_3$  or  $n_2 : n_3 = A_2 : A_3$ .

Another estimator of the ratio  $\phi$  is

$$(64) \quad f = P_1 f_1 + P_2 f_2 + P_3 f_3,$$

wherein  $P_i = B_i/B$  and  $f_i = \bar{x}_i/\bar{y}_i$ . This estimator is sometimes preferred when  $f_i$  varies greatly from stratum to stratum, and when there can be no trouble with small denominators. The allocation of sample for this estimator is, for practical purposes, the same as in eq. (63) (Cochran [1953] 1963, p. 175; Hansen et al. 1953, vol. 1, p. 209).



*Sequential adjustment of size of sample.* It is sometimes possible, when decision on the size of sample is difficult, or when time is short, to break the sample in advance into two portions, 1 and 2, each being a valid sample of the whole frame. Portion 1 is definitely to be carried through to completion, but portion 2 will be used only if required. This may be called a two-stage sequential method. It is practicable where the investigation is to be carried out by a small number of experts that will stay on the job as long as necessary but not where a field force must be engaged in advance for a definite period.

*Modifications for differing costs.* If investigating a sampling unit in a particular stratum is three or more times as costly as the average investigation, it may be wise to decrease the sample in the costly stratum and to build up the sample in other strata (Deming 1960, p. 303).

*Considerations for planning.* In order to plan a stratified sample, certain assumptions are necessary. Fair approximations to the relevant ratios, such as  $\sigma_w : \sigma$ ,  $\sigma_{w_0} : \sigma$ ,  $\sigma_{w_1} : \sigma$ ,  $\sqrt{c_1} : c_2$ , will provide excellent allocation. On the other hand, bad approximations to these ratios, or failure to use theory at all, can lead to serious losses.

The required good approximations to these ratios may come from prior experience, or from probing the knowledge of experts in the subject matter. For example, the distribution of intelligence quotients in the stratum between 90 and 110, if rectangular, would provide  $\sigma^2 = (110 - 90)^2/12$ , or 33, whence  $\sigma = 5.7$ . Other shapes have other variances, but shape is fortunately not critical (Deming 1950, p. 262; 1960, p. 260). A stratum with very high values should be set off for special treatment and possibly sampled 100 per cent.

*Stratification for analytic studies.* As mentioned earlier, the aim in an analytic study is to detect differences between classes or to measure the effects of different treatments.

The general formula for the variance of the difference between two means,  $\bar{x}_A$  and  $\bar{x}_B$ , derived from independent samples of sizes  $n_A$  and  $n_B$  drawn by random numbers singly and without stratification from, for example, two groups of patients, A and B, is

$$(65) \quad \text{var}(\bar{x}_A - \bar{x}_B) = \sigma_A^2/n_A + \sigma_B^2/n_B,$$

wherein  $\sigma_A^2$  and  $\sigma_B^2$  are the respective variances between the patients within the two groups.

For such analytic studies the optimum allocation of skill and effort is found by setting

$$(66) \quad \frac{n_A}{n_B} = \frac{\sigma_A}{\sigma_B} \sqrt{\frac{c_B}{c_A}},$$

wherein  $c_A$  and  $c_B$  are the costs per case. Note that the sizes of the groups do not enter into this formula and that it is different from the optimum allocation in enumerative problems.

In many analytic studies  $\sigma_A$  and  $\sigma_B$  will be about equal, and so will the costs  $c_A$  and  $c_B$ . In such circumstances, the best allocation is

$$(67) \quad n_{.1} = n_{.2}.$$

### Regression estimators

We have already seen reduction in variance resulting from use of prior or supplementary knowledge concerning the frame. Use of prior knowledge of  $N$  to form the estimator  $X = N\bar{x}$  is an instance. Prior knowledge of  $B$  to form the ratio estimator is another instance. This section, on *regression estimators*, describes other ways to use prior or supplementary knowledge concerning the frame. Regression estimators include the simple estimator,  $\bar{x}$ , and the ratio estimator,  $fb$ , as special cases, but they also include many other estimators, some of them highly useful. Like the ratio estimator of a total, these additional estimators are applicable only if independent and fairly reliable information is available about the  $y$ -population in the frame. Any estimator that takes advantage of supplementary information may have considerable advantage over the simple estimator,  $\bar{x}$ , if the correlation,  $\rho$ , between  $x_i$  and  $y_i$  is high, but this condition is not in itself sufficient.

*Specific forms of regression estimators.* Assume simple random selection and write the regression estimator in the form

$$(68) \quad \bar{x}_i = \bar{x} + m_i(b - \bar{y}),$$

wherein  $b$  is known independently from some source such as a census. The subscript  $i$  on  $\bar{x}_i$  here differentiates the several specific forms of regression estimators.

Regression estimators are closely allied with the analysis of covariance [see LINEAR HYPOTHESES, article on ANALYSIS OF VARIANCE]. The four cases to be considered here are taken largely from Hansen, Hurwitz, and Madow (1953).

*Simple estimator.* If  $m_i$  is taken as zero, the regression estimator obtained is  $\bar{x}_i = \bar{x}$ , seen earlier. This procedure makes no use of supplemental information. Under the assumption that  $N$  is large relative to  $n$ , the variance of this estimator is

$$(69a) \quad \text{var} \bar{x}_i = \sigma_x^2;$$

likewise,

$$(69b) \quad \text{var} \bar{y}_i = \sigma_y^2.$$

*Difference estimator.* The estimator  $\bar{x}_2$ , often called the difference estimator, is practicable if

Table 4 — Rel-variances when estimator of  $b$  is subject to sampling error

Estimator	Case I: sample of size $n$ is drawn as a subsample of $n'$	Case II: samples of size $n$ and $n'$ are independent
	Same as in Case I	Same as in Case I
$\bar{x}_1$	$C_y^2$	$C_y^2$
$\bar{x}_2$	$C_y^2[1 - \rho^2(1 - e^2)(1 - n/n')]$	$C_y^2[1 - \rho^2(1 - e^2) + \rho^2(1 + e)^2 n/n']$
$\bar{x}_3$	$C_y^2[1 - \rho^2(1 - n/n')]$	Same as in Case I
$\bar{x}_4$	$C_y^2 - (2\rho C_x C_y - C_y^2)(1 - n/n')$	$C_y^2 - (2\rho C_x C_y - C_y^2)(1 - n/n') + (2C_y n/n')(C_y - \rho C_x)$

prior knowledge (such as prior surveys of a related type) provides a rough approximation to the regression coefficient  $\beta = \rho\sigma_x/\sigma_y$ . This estimator is

$$(70) \quad \bar{x}_2 = \bar{x} + m_2(b - \bar{y}),$$

where  $m_2$  is any approximate slope not derived from the sample under consideration. The variance of  $\bar{x}_2$  is

$$(71) \quad \begin{aligned} \text{var } \bar{x}_2 &= \sigma_x^2(1 - \rho^2) + \sigma_y^2(m_2 - \beta)^2 \\ &= \sigma_x^2(1 - \rho^2 + \rho^2 e^2), \end{aligned}$$

where  $\beta = \rho\sigma_x/\sigma_y$  and  $e = (m_2 - \beta)/\beta$ . (Note that  $\rho e = (\sigma_y/\sigma_x)(m_2 - \beta)$  even if  $\rho = 0$ .)

*Least squares regression estimator.* If  $m_1$  is chosen as

$$(72) \quad m_1 = \frac{\sigma_{xy}}{\sigma_x^2} = \rho \frac{\sigma_y}{\sigma_x},$$

then the equation

$$(73) \quad \bar{x}_3 = \bar{x} + m_3(b - \bar{y})$$

gives the so-called least squares regression estimator. The variance of this estimator is

$$(74) \quad \text{var } \bar{x}_3 = \sigma_x^2(1 - \rho^2) + R.$$

Here  $R$  is a remainder in the Taylor series involving  $1/n^2$  and higher powers; this remainder will be negligible if  $n$  is large.

*Ratio estimator.* If  $m_4$  is chosen as

$$(75) \quad m_4 = \bar{x}/\bar{y} = f,$$

the ratio estimator is

$$(76) \quad \bar{x}_4 = f\bar{b}.$$

The variance of  $\bar{x}_4$  is

$$(77) \quad \text{var } \bar{x}_4 = \sigma_x^2(1 + C_y^2/C_x^2 - 2\rho C_y/C_x) + R',$$

$R'$  being another remainder. For large  $n$  and for  $C_x \approx C_y$ ,

$$(78) \quad \text{var } \bar{x}_4 \approx 2\sigma_x^2(1 - \rho).$$

It follows that for large  $n$  and for  $m_2 \approx \beta$  and  $C_x \approx C_y$ ,

$$(79) \quad \frac{\text{var } \bar{x}_4}{\text{var } \bar{x}_2} \approx \frac{2}{1 + \rho}.$$

*Comparison of regression estimators.* If the correlation,  $\rho$ , between  $x_i$  and  $y_i$  is moderate or high,

but the line of regression of  $x$  on  $y$  misses the origin by a wide margin, then the estimator  $\bar{x}_3$  will show substantial advantages over  $\bar{x}_4$  and  $\bar{x}_1$ . If the  $y$ -variate shows relatively wide spread (that is, if  $C_y$  is much greater than  $C_x$ ), the ratio estimator  $\bar{x}_4$  may be far less precise than the simple estimator  $\bar{x}_1 = \bar{x}$ , even when  $\rho$  is high, especially if the line of regression misses the origin by a wide margin. On the other hand, if the line of regression passes through the origin ( $\rho C_x = C_y$ ), or nearly through it,  $\bar{x}_3$  and  $\bar{x}_4$  will have about the same variance, but  $\bar{x}_1$  may be much easier to compute.

*Estimator of  $b$  subject to sampling error.* It often happens that the  $y$ -population per sampling unit is not known with the reliability of a census but comes instead from another and bigger sample. This circumstance introduces additional terms into the variances. Let  $n$  be the size of the present sample and  $n'$  the size of the sample that provides the estimate of  $b$ . We suppose that the variance of this estimate of  $b$  is  $\sigma_b^2/n'$ . The resulting variances of the regression estimators are shown in Table 4.

W. EDWARDS DEMING

#### BIBLIOGRAPHY

- BOWLEY, ARTHUR L. (1901) 1937 *Elements of Statistics*. 6th ed. New York: Scribner; London: King.
- CHEVRY, GABRIEL 1949 Control of a General Census by Means of an Area Sampling Method. *Journal of the American Statistical Association* 44:373-379.
- COCHRAN, WILLIAM G. 1946 Relative Accuracy of Systematic and Stratified Random Samples for a Certain Class of Populations. *Annals of Mathematical Statistics* 17:164-177.
- COCHRAN, WILLIAM G. (1953) 1963 *Sampling Techniques*. 2d ed. New York: Wiley.
- DALENIUS, TORE 1962 Recent Advances in Sample Survey Theory and Methods. *Annals of Mathematical Statistics* 33:325-349.
- DEMING, W. EDWARDS (1943) 1964 *Statistical Adjustment of Data*. New York: Dover.
- DEMING, W. EDWARDS 1950 *Some Theory of Sampling*. New York: Wiley.
- DEMING, W. EDWARDS 1960 *Sample Design in Business Research*. New York: Wiley.
- DEMING, W. EDWARDS; and KEYFITZ, NATHAN 1965 Theory of Surveys to Estimate Total Population. Volume 3, pages 141-144 in *World Population Conference, Belgrade, August 30-September 10, 1965, Proceedings*. New York: United Nations.



- FISHER, R. A. (1922) 1950 *On the Mathematical Foundations of Theoretical Statistics*. Pages 10.307a-10.368 in R. A. Fisher, *Contributions to Mathematical Statistics*. New York: Wiley. → First published in Volume 222 of the *Philosophical Transactions*, Series A, of the Royal Society of London.
- FISHER, R. A. (1956) 1959 *Statistical Methods and Scientific Inference*. 2d ed., rev. New York: Hafner.
- GAUSS, CARL FRIEDRICH 1823 *Theoria combinationis observationum erroribus minimis obnoxiae*. Göttingen (Germany): Dieterich. → A French translation was published in Gauss' *Méthode des moindres carrés* (1855). An English translation of the French was prepared as *Gauss's Work (1803-1826) on the Theory of Least Squares*, by Hale F. Trotter; Statistical Techniques Research Group, Technical Report, No. 5, Princeton Univ., 1957.
- HANSEN, MORRIS H.; and HURWITZ, WILLIAM N. 1943 *On the Theory of Sampling From Finite Populations*. *Annals of Mathematical Statistics* 14:333-362.
- HANSEN, MORRIS H.; and HURWITZ, WILLIAM N. 1946 *The Problem of Non-response in Sample-surveys*. *Journal of the American Statistical Association* 41: 517-529.
- HANSEN, MORRIS H.; HURWITZ, WILLIAM N.; and MADOW, WILLIAM G. 1953 *Sample Survey Methods and Theory*. 2 vols. New York: Wiley.
- KISH, LESLIE 1965 *Survey Sampling*. New York: Wiley. → A list of errata is available from the author.
- KOLLER, SIEGFRIED 1960 *Aussenhandelsstatistik: Untersuchungen zur Anwendung des Stichprobenverfahrens*. Pages 361-370 in Germany (Federal Republic), Statistisches Bundesamt, *Stichproben in der amtlichen Statistik*. Stuttgart (Germany): Kohlhammer.
- LEVEN, MAURICE 1932 *The Income of Physicians: An Economic and Statistical Analysis*. Univ. of Chicago Press.
- MAHALANOBIS, P. C. 1944 *On Large-scale Sample Surveys*. Royal Society of London, *Philosophical Transactions* Series B 231:329-451.
- MAHALANOBIS, P. C. 1946 *Recent Experiments in Statistical Sampling in the Indian Statistical Institute*. *Journal of the Royal Statistical Society Series A* 109: 326-378. → Contains eight pages of discussion.
- MOSE, C. A. 1949 *The Use of Sampling in Great Britain*. *Journal of the American Statistical Association* 44:231-259.
- NEYMAN, JERZY 1934 *On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection*. *Journal of the Royal Statistical Society Series A* 97:558-606.
- NEYMAN, JERZY 1938 *Contribution to the Theory of Sampling Human Populations*. *Journal of the American Statistical Association* 33:101-116. → See especially equation 49 on page 110.
- QUENOUILLE, M. H. 1959 *Rapid Statistical Calculations*. London: Griffin; New York: Hafner. → Pages 5-7 show estimators of the standard deviation by use of the range.
- SATTERTHWAITE, F. E. 1946 *An Approximate Distribution of Estimates of Variance Components*. *Biometrics* 2:110-114.
- SHEWHART, WALTER A. 1939 *Statistical Method From the Viewpoint of Quality Control*. Washington: U.S. Department of Agriculture, Graduate School.
- STEPHAN, FREDERICK F. 1936 *Practical Problems of Sampling Procedures*. *American Sociological Review* 1:569-580.
- STEPHAN, FREDERICK F. 1948 *History of the Uses of Modern Sampling Procedures*. *Journal of the American Statistical Association* 43:12-39.
- STEPHAN, FREDERICK F.; and MCCARTHY, PHILIP J. 1958 *Sampling Opinions: An Analysis of Survey Procedure*. New York: Wiley.
- STUART, ALAN 1962 *Basic Ideas of Scientific Sampling*. London: Griffin; New York: Hafner.
- SYMPOSIUM ON CONTRIBUTIONS OF GENETICS TO EPIDEMIOLOGIC STUDIES OF CHRONIC DISEASES, ANN ARBOR, MICHIGAN, 1963 1965 *Genetics and the Epidemiology of Chronic Diseases*. U.S. Public Health Service, Publication No. 1163. Washington: Government Printing Office. → See especially "Selection techniques for rare traits," by Leslie Kish, pages 165-176.
- YATES, FRANK (1949) 1960 *Sampling Methods for Censuses and Surveys*. 3d ed., rev. & enl. New York: Hafner. → Earlier editions were also published by Griffin.

## II

## NONPROBABILITY SAMPLING

Nonprobability sampling refers to the selection of sampling units for a statistical study according to some criterion other than a probability mechanism. In contrast to probability sampling, nonprobability sampling has two major disadvantages: (1) the possibility of bias in the selection of sampling units and (2) the impossibility of calculating sampling error from the sample.

The attitude toward sampling held by most social scientists has come full circle: whereas it was once necessary to argue strenuously the case for deliberately using a probability mechanism in the selection of material for study, that case is now so widely accepted that any nonprobabilistic sampling procedure is regarded with suspicion. It will be argued here that such suspicion is generally well founded but that it is not to be equated with condemnation, since there are occasions when nonprobability sampling is the only procedure open to the investigator.

The issues involved in nonprobability sampling may be sharply focused by considering first the status of the inferences made by an archeologist. He can usually expect to find only a fraction of the material that he would like in order to examine the causes and motives of past events, and that fraction will generally be small for remote periods. Every object that he would judge relevant is a potential witness in his court of inquiry, but many of them have long since been destroyed or lost. Yet there is no contradicting the fact that some strongly convincing arguments can be made, even about the remote past. Does it follow, then, that probability sampling is inessential to any social scientist's inferences about the present?

As soon as it is drawn, this analogy is seen to be false. The archeologist usually has no choice;

he must, in the court metaphor, accept what witnesses offer themselves. The situation of another social scientist is different: he is rather in the position of having too many potential witnesses, among whom he must choose on grounds of cost and time. How is he to choose? He cannot fairly exclude those potential witnesses whose evidence may not be to his taste, for if he did, he would implicitly abandon his judicial position and become merely an advocate. (The fact that some have apparently undergone this transformation is not good reason for others to follow them.) If some sources of evidence must be chosen in preference to others, they can only conscientiously be chosen in some way unrelated to the nature of their evidence: they must be selected by a probability mechanism.

The value of probability sampling is, therefore, not that it ensures employment for statisticians, but that it provides a guarantee of freedom from selection bias on the part of the investigator, bias that may be strong even if quite unconsciously produced (Yates [1949] 1960, chapter 2). The absence of selection bias is the first advantage of probability sampling.

Nonprobability sampling always has the characteristic that the sampling procedure is ill-defined. One literally cannot know what chance of selection any individual in the population has had. It is this lack of precise definition that leads directly to the second disadvantage of nonprobability sampling: there is no way of estimating the sampling error of a nonprobability sample from the sample itself, while such estimation is typically possible when probability sampling is used.

Despite what has just been said, nonprobability samples will always be important in the social sciences. Confronted with a remote and hostile tribe, the anthropologist will not be inclined to imperil his foothold of hospitality and cooperation by selecting informants at random. In advanced societies, selected individuals or institutions may refuse to cooperate in an inquiry and usually cannot be compelled to do so. Even in such circumstances, however, at least an attempt at a probability sample should always be the investigator's object. He may fail to achieve it but cannot be worse off than if he had not tried at all. The unshakable optimism of all candidates for political office must be partly attributed to reports from supporters who have largely been sampling the faithful.

Nonprobability samples can arise in several ways: through the investigator's ignoring or denying the force of the above considerations; through his recognizing their force but claiming that the

objectives of probability sampling can just as well be achieved by substitute methods; or finally, through his lack of success in attempts to achieve probability samples. About the first category nothing more will be said here, but a more detailed examination of the other two will be made.

**Representative methods.** Most of the substitutes suggested for probability sampling have in common some attempt to make the sample "representative" of the population from which it was selected, for example, by arranging that the proportions of men and women be the same in both or (more rarely) that the average income be the same in both. In practice, the most frequent substitute is quota sampling (Moser 1952), in which distributions by sex, age, "class," and sometimes additional characteristics are equalized between sample and population and the interviewers are otherwise free to select cases. Such representative methods are to be sharply distinguished from stratified probability sampling, where the population is divided into groups prior to sampling and each group is sampled separately by probability methods.

There are several distinct criticisms to be made of representative procedures. In the first place, the agreement of even quite a large number of averages or percentages between sample and population in no way guarantees high accuracy in other respects, and very large biases are still possible (Neyman 1934). Second, the lack of definition in the sampling procedure is not corrected by the imposition of such agreements; sampling error is still not calculable.

The most important criticism, however, is a paradoxical one. There is no such thing as a "representative," "unbiased," "fair," or otherwise "acceptable" sample: such adjectives are strictly applicable to the sampling process that produces the sample, and their application to the sample itself is at best a piece of verbal shorthand and at worst a sign of muddled thinking. A sample can be judged only in relation to the process that produced it. The central concepts of selection bias and sampling error have no meaning except in this context (Stuart 1962). Thus, an ill-defined sampling process, such as is involved in nonprobability samples, can only produce a sample with ill-defined properties of bias and sampling error.

**Embedding in a probability framework.** The statement that there is no way of calculating sampling error for nonprobability samples is always true but can be made irrelevant by the device of embedding a nonprobability sampling procedure within a higher-order probability framework. Suppose that a nonprobability sampling procedure is proposed in which 1,000 families are to be inter-



viewed in a certain area and that ten interviewers are to be used. If each interviewer is independently given an assignment to interview 100 families by identical methods, there will be not one, but ten nonprobability samples, and the variation among the results of these ten samples may be used to estimate the sampling error attached to a sample of 100 families and thence, by a natural extension, to estimate the error for a sample of 1,000 families. Since the sample is designed as ten equivalent independent subsamples, in effect the sampling is of ten members from the population of all possible samples obtainable by that method with those interviewers.

In practice, it may be difficult or practically impossible to achieve independent subsamples in some contexts. For example, in the construction of index numbers the choice of commodity items is usually made by a panel of experts, so that the items chosen constitute a nonprobability sample. If an attempt is made to measure the variation between the judgments of several experts at the same level of expertise, interaction phenomena that make the measurement difficult are immediately encountered. For example, the experts are likely to have consulted one another; they are likely to know the criteria of selection used by their peers; they are bound to meet in the course of examining files containing background information, and so on. Thus, although independent judgments are in principle possible, the practical problem of arranging for independence is very difficult. Now, almost certainly, in the simple interviewing example, differences between the interviewers will affect the results of the survey (through personal characteristics, differences in thoroughness, etc.), and such differences will inflate the sampling error when it is estimated as above. If it were important to eliminate differences between interviewers from the sampling error, a slight modification in sampling design would suffice. Each interviewer must be asked to carry out at least two independent sample assignments, say four assignments of 25 families each. It will then be possible, by exactly the same argument as before, to estimate variability "within interviewers" only, thus excluding variability "among interviewers" from sampling error estimates. Further improvements in design are also possible. [See INDEX NUMBERS, *article on SAMPLING*.]

It will be seen that nonprobability sampling treated in this way utilizes the theory of experiment design, and, as always in that subject, care is necessary to ensure that assignments are allocated at random to interviewers. An incidental

advantage of such experiment designs, which is often of greater practical importance than the original purpose of estimating sampling error, is that they make it possible to see whether interviewers are varying so much in performance that further training or other action is called for.

Such evidence as is available from designs of this sort indicates (Moser & Stuart 1953; Stephan & McCarthy 1958) that the sampling errors of quota sampling are considerably larger than those of comparable probability sampling. (If sampling is multistage, the inflation will of course only apply to the stage or stages at which quota sampling is used—this is commonly the final stage only.) Quota sampling can therefore only be justified, if at all, on grounds of reduction of costs, but it is doubtful whether, in the light of current cost structures, this factor is large enough to offset the excess sampling error of quota sampling. Some crude numerical guide is contained in the introduction to the tables by Stuart (1963).

Of course, experiment designs of this sort may be used with any ordinary probability sampling scheme (Mahalanobis 1946). The point here is that experiment designs rescue nonprobability sampling schemes from one of their worst deficiencies, the impossibility of calculating sampling error estimates, which does not hold for probability sampling. Nothing, however, can rescue nonprobability sampling from the ever-present danger of selection biases, a danger particularly acute for those forms (like quota sampling) that allow considerable freedom to interviewers to select cases.

Taken in conjunction with the sampling error and cost considerations, the bias danger suggests that, whatever may have been the case when it originated, quota sampling is now, in terms of reliable information delivered per unit cost, uneconomical as well as anachronistic.

**Incompletely achieved probability samples.** It has been noted that the major unremovable drawback of nonprobability samples is the danger of bias in the selection procedure. Inevitably, this danger must also arise when an intended probability sample is, through lack of cooperation, through inaccurate fieldwork, or through accidentally missing records, reduced to a fraction of the number of cases originally selected. The selected sample is, for one or another of these reasons, incompletely achieved, and the sampler's duty is to survey the damage done to his original plan and try to assess the importance of the danger to his purposes. Strictly speaking, an incompletely achieved probability sample ceases to be a probability sample, although it usually continues to be

called one. Intentions are not the same as achievement, and the sampler must see whether the courtesy title is deserved.

Before discussing what can be done in such situations, it is useful to compare the reaction of the probability sampler confronted by an incompletely achieved sample with that of the nonprobability sampler in the same position. It is an empirical fact that the probability sampler is usually much more worried about the incomplete sample than is the nonprobability sampler, because the stringency of a probability sampling scheme draws much greater attention to any weaknesses in the achieved sample. By contrast, the nonprobability sampler usually finds it easier to bring his sample numbers up to the required level. His sampling scheme is less stringent and can more easily be patched up. Advocates of nonprobability sampling (principally quota sampling) have been known to categorize this as an advantage for their method, but we shall take the contrary view. It is because nonprobability sampling is so ill-defined that its definition can be extended to cover incomplete sample fulfillment. In the kingdom of the one-eyed, the blind man is less easily seen. Perhaps the least appreciated, because it is the most troublesome, feature of probability sampling is the urge for sample completeness that it imposes upon its practitioners.

Faced with a seriously incomplete sample, the probability sampler will usually redouble his efforts, but however hard he tries, he will ultimately have to face the fact that some part of his sample is not to be completed. The characteristic rates of noncompletion vary greatly with types of sample. For probability samples of individuals in the United Kingdom, it is possible to achieve completion rates of 90 per cent, although 80 per cent is nearer the general average. For probability samples of households in United Kingdom government household-budget inquiries, the completion rates range around 70 per cent. Few surveys involve a more onerous burden upon the respondent than the keeping of a detailed household budget, so it is probably reasonable to state that completion rates in the United Kingdom should never, with careful fieldwork, fall below about 70 per cent. The range of noncompletion rates likely to be encountered is thus of the order of 10 to 30 per cent. For many purposes, the sample results will not be seriously biased by even a noncompletion rate of one-quarter. For example, if 60 per cent of all households in the completed three-quarters of the sample display a certain characteristic, the non-completed quarter of the sample would need to

have under 40 per cent or over 80 per cent with that characteristic for the bias in the completed sample to exceed 5 per cent.

The crucial question to be examined in assessing the likely magnitude of bias due to incompleteness is whether the causes of incompleteness are related to the questions of interest. In a study of women's cosmetics usage, it would be running a serious risk to ignore a very low completion rate for teen-agers, who are particularly heavy users of cosmetics; if the completed sample showed a very low percentage of teen-agers compared with the selected sample (or known recent population data), special efforts would have to be made to increase that percentage. But if the completed sample showed an excess of women aged 45-55 and a shortage of women over 65, this would be unlikely to exert a significant influence on cosmetics-usage findings. However, if the subject studied were demand for domestic help, it would be the shortage of the elderly that would be the threat. Data from published population figures or from specially undertaken supplementary probability samples can often be used to check and, with appropriate methods of statistical analysis, to improve estimates suspected of bias, whether it is due to sample incompleteness or not (Moser & Stuart 1953; United Nations 1960, sec. 13). The simple examples just given show that such check data can only be used in conjunction with the experienced judgment of the investigator. If extreme imbalances of the completed sample are revealed, they are just cause for suspicion of the sample results, but the converse does not hold: no finite amount of external checking of this kind can ever fully validate a nonprobability sample, since the crucial variable to check may be overlooked or not checkable. There is no completely satisfactory substitute for a fully achieved probability sample. Any shortfall from this status is a potential threat to the inferences drawn from the sample. The investigator may be able to judge in some cases that the threat is not likely to be serious, but he can do no more than this. Insofar as a probability sample is fully achieved, it obviates the need to make such judgments.

To return to the first analogy, it can now be seen that the archeologist is in a situation closely resembling that of the investigator with an incomplete nonprobability sample. Another social scientist can usually improve on this situation at least to the extent of starting with a probability sample of his material. The incompleteness of his sample may on occasion compel him to make judgments



of an inconclusive kind about the quality of his sample, but this is not a valid reason for extending the scope of the inconclusive judgment to the whole sampling procedure.

ALAN STUART

[For methods of accomplishing randomization, see EXPERIMENTAL DESIGN, article on THE DESIGN OF EXPERIMENTS; RANDOM NUMBERS. Also related are ERRORS, article on NONSAMPLING ERRORS; EXPERIMENTAL DESIGN, article on QUASI-EXPERIMENTAL DESIGN; INDEX NUMBERS, article on SAMPLING.]

#### BIBLIOGRAPHY

- MAHALANOBIS, P. C. 1946 Recent Experiments in Statistical Sampling in the Indian Statistical Institute. *Journal of the Royal Statistical Society Series A* 109: 326-378. → Contains eight pages of discussion.
- MOSER, CLAUS A. 1952 Quota Sampling. *Journal of the Royal Statistical Society Series A* 115:411-423.
- MOSER, CLAUS A. 1958 *Survey Methods in Social Investigation*. New York: Macmillan.
- MOSER, CLAUS A.; and STUART, ALAN 1953 An Experimental Study of Quota Sampling. *Journal of the Royal Statistical Society Series A* 116:349-405. → Contains 11 pages of discussion.
- NEYMAN, JERZY 1934 On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection. *Journal of the Royal Statistical Society Series A* 97: 558-606.
- STEPHAN, FREDERICK F.; and MCCARTHY, PHILIP J. 1958 *Sampling Opinions: An Analysis of Survey Procedure*. New York: Wiley.
- STUART, ALAN 1962 *Basic Ideas of Scientific Sampling*. London: Griffin; New York: Hafner.
- STUART, ALAN 1963 Standard Errors for Percentages. *Applied Statistics* 12:87-101.
- UNITED NATIONS, DEPARTMENT OF ECONOMIC AND SOCIAL AFFAIRS 1960 *A Short Manual on Sampling*. Volume 1: Elements of Sample Survey Theory. United Nations, Statistical Office, Studies in Methods, Series F, No. 9. New York: United Nations.
- YATES, FRANK (1949) 1960 *Sampling Methods for Censuses and Surveys*. 3d ed., rev. & enl. London: Griffin; New York: Hafner.

#### SAMPLING

See SAMPLE SURVEYS.

International  
Encyclopedia of the  
SOCIAL  
SCIENCES



## Associate Editors

Heinz Eulau, *Political Science*  
Lloyd A. Fallers, *Anthropology*  
William H. Kruskal, *Statistics*  
Gardner Lindzey, *Psychology*  
Albert Rees, *Economics*  
Albert J. Reiss, Jr., *Sociology*  
Edward Shils, *Social Thought*

## Special Editors

Elinor G. Barber, *Biographies*  
John G. Darley, *Applied Psychology*  
Bert F. Hoselitz, *Economic Development*  
Clifford T. Morgan, *Experimental Psychology*  
Robert H. Strotz, *Econometrics*

## Editorial Staff

Marjorie A. Bassett, *Economics*  
P. G. Bock, *Political Science*  
Robert M. Coen, *Econometrics*  
J. M. B. Edwards, *Sociology*  
David S. Gochman, *Psychology*  
George Lowy, *Bibliographies*  
Judith M. Tanur, *Statistics*  
Judith M. Treistman, *Anthropology*

Alvin Johnson

HONORARY EDITOR

W. Allen Wallis

CHAIRMAN, EDITORIAL ADVISORY BOARD

# International Encyclopedia of the SOCIAL SCIENCES

DAVID L. SILLS EDITOR

VOLUME 14

The Macmillan Company & The Free Press, New York  
COLLIER-MACMILLAN PUBLISHERS, LONDON



COPYRIGHT © 1968 BY CROWELL COLLIER AND MACMILLAN, INC.

ALL RIGHTS RESERVED UNDER THE INTERNATIONAL COPYRIGHT UNION,  
THE INTER-AMERICAN COPYRIGHT UNION, AND UNDER THE PAN-AMERICAN  
COPYRIGHT CONVENTIONS.

NO PART OF THIS BOOK MAY BE REPRODUCED OR TRANSMITTED IN  
ANY FORM OR BY ANY MEANS, ELECTRONIC OR MECHANICAL, INCLUDING  
PHOTOCOPYING, RECORDING, OR BY ANY INFORMATION STORAGE AND  
RETRIEVAL SYSTEM, WITHOUT PERMISSION IN WRITING FROM  
CROWELL COLLIER AND MACMILLAN, INC.

MANUFACTURED IN THE UNITED STATES OF AMERICA

REPRINT EDITION 1972

International  
Encyclopedia of the  
SOCIAL  
SCIENCES





# S

[ CONTINUED ]

## SANCTIONS

Social control involves the process by which a group perpetuates its culture and system of social relations. In what has become an anthropological *locus classicus*, Radcliffe-Brown (1934) defined the sanction as a social reaction to a mode of behavior that was thereby approved or disapproved. In his view the concept is of primary significance to sociology because sanctions provide an effective instrument for regulating conduct in conformity with social usage. This formulation, however, seems to obscure a number of important points. Radcliffe-Brown's argument clearly implies that the sanction, at least in one of its aspects, has a deterrent function. But if this object is to be achieved, logic demands that the sanction must precede the reaction: the sanction must already be in operation when a breach of norm occurs. It seems preferable therefore to define sanctions in terms of a promise of reward for fulfillment of the norms of behavior that are socially recognized and accepted, or in the liability to suffer the consequences that attend their breach. According to this view, the sanction is thus characterized by a quality of latency that is expressed most clearly, in the case of negative sanctions, in a general belief in the continuing likelihood of the execution of the threat. This implies two further properties of the sanction. A sanction cannot be arbitrary or *ad hoc*: its existence must be generally known, and it must be regular in its incidence. In fulfilling these conditions, the sanction serves not merely to deter; it also indicates the appropriate steps that may be taken to redress the situation when a norm has been infringed or to express approbation when it has been overfulfilled.

The sanction is, of course, a concept of funda-

mental importance in jurisprudence and has played a paramount role in different legal theories. But there seems no good reason why its use should be restricted in this way. The legal sanction, manifested in the form of the penalties, remedies, and modes of redress of the law, represents only one particular means of enforcing conformity to norms and restating their validity when they are breached. Each group and subgroup within a society tends to develop its own distinctive pattern of usages and the means of maintaining them without necessary recourse to the municipal law. Sanctions therefore come to operate within every conceivable set of group relationships: they include not only the organized sanctions of the law but also the gossip of neighbors or the customs regulating norms of production that are spontaneously generated among workers on the factory floor. In small-scale communities, or within segments of a large-scale society, informal sanctions may become more drastic than the penalties provided for in the legal code. Thus the concept of sanctions is not merely of jurisprudential interest; it also has immediate sociological relevance to the analysis of the problem of social control.

It will be apparent then that social control is achieved through no single mechanism. On the contrary, different principles and processes will frequently be found working together in any given situation. Accordingly, it is only by specifying the character of the sanctions present in the situation that one can isolate the different elements for analytical purposes and give them their proper weighting. The institution of law represents, of course, one major instrument of social control, and it is most usually marked off from other such agencies by virtue of the fact that its sanctions are



imposed by politically organized society. As Seagle (1941) has observed, for example, it is the fact that the sanction is applied exclusively by organized political government that distinguishes the sphere of law from the spheres of religion, morals, and custom.

**The mechanism of sanction.** The differentiation of the instruments of social control in stateless and state societies and in peasant communities and bands of hunting and gathering people has provided ground for controversy between jurists and anthropologists and among anthropologists themselves. Although their discussions have not always clarified the nature of law in different types of societies, they have been fruitful in opening up the broader sociological problem of social control. Malinowski's position as a pioneer in this field remains undisputed. Working in an island society that lacked any of the familiar machinery for the enactment, administration, and enforcement of law, Malinowski was led to seek elsewhere for the sources of social cohesion. He came to the conclusion that there was a class of rules "too practical to be backed up by religious sanctions, too burdensome to be left to mere goodwill." This constituted the Trobriand "civil law," the binding character of which lay in the sanctions of reciprocity and systematic incidence (Malinowski 1926). Later critics have pointed out that reciprocity is a concept of such wide generality that it cannot usefully serve to delineate the sphere of law. A curious feature of Malinowski's thinking on these matters is that, despite his emphasis on conflict, he appears not to have appreciated that just as the sanction implies the potential breach of norm, so much of law has to be concerned with what may be called the "pathology" of social relations. The emphasis on reciprocity has been important in deepening our understanding of why rules of behavior are accepted in so many areas of social life without the need to resort to formal mechanisms of redress. The question that remains unanswered is what happens when reciprocity breaks down and breach actually occurs. The point had been raised often enough, but it was left to Llewellyn and Hoebel (1941) in their now classic study of the Cheyenne to repair the defect and put the anthropological study of legal sanctions back on its proper path. In their succinct and expressive phrase, "the law has teeth": their procedure was to show that, even in societies lacking instituted juridical bodies, the investigation of instances of "hitch," dispute, grievance, and trouble led directly into the way in which physical coercion was organized.

Pospisil (1958), on the basis of his work among

Kapauku Papuans, has suggested that too great an emphasis has been placed upon physical force in defining the character of the legal sanction, pointing out that cultures exist where physical sanctions are practically lacking. This introduces an unnecessary confusion. The crucial question is not the actual mode of coercion employed but recognition of the legitimate authority in which is vested the power of coercion. Such authority is certainly recognized among the Kapauku, although coercion may frequently be achieved by other than purely physical means. Where, in a society lacking developed forensic institutions, rules of conduct may be enforced through the powers of coercion vested in a legitimate authority, I propose to speak of jural rather than legal sanctions, indicating that here we are in the realm of *ius*, but not of *lex*.

**Jural sanctions.** Once the criterion of legitimate coercion is adopted, it becomes apparent that primitive societies, lacking the formal machinery of the law, may yet display a remarkable range of sanctions that can be distinguished as jural. It has sometimes been claimed of many of the simplest societies that their social life is dominated by notions of the supernatural: there is a paucity of rules of law that is the reflex of the comprehensiveness of taboo. Such views, however, do not do justice to the facts. Thus it is reported of the Eskimo, for example, that persistent breach of taboo may be met by a shaman's decree of banishment. More strikingly, it is also reported that recidivist homicides may be dispatched by an individual acting at the request of the local community or with its compliance; this individual does not incur vengeance. Another interesting type of jural sanction was reported by Radcliffe-Brown in the Andaman Islands (1922). There, aggrieved persons were apparently able to vent their spleen by destroying any property they could lay their hands on, but unfortunately there are a number of points touching the operation of the sanction which the sources do not make very clear. It is possible that the Andamanese practice was somewhat similar to the custom of *hamara* reported from the Gazelle Peninsula of New Britain. *Kamara* took a number of variant forms, but essentially it was a device whereby a person who had suffered injury could damage or distrain the property of someone of acknowledged influence within the community. This individual was expected then to bring pressure on the offender to make retribution.

But in general, within the range of primitive societies, the most common mode of redress for wrong is found to rest on the principle of self-help. In these circumstances, the legal initiative lies in-

variably with the individual or group against whom a wrong has been committed. Yet the exacting of vengeance is not an arbitrary procedure. It has to be carried out according to certain rules and is undertaken only when morally supported by the community or because, indeed, it has been pressed upon individuals or groups by expressions of public approval or disapproval.

The principle of self-help involves a point of considerable importance for the discussion of sanctions. This is what Nadel (1947) has called the social range of offenses, where the kind of sanction provoked depends upon the nature of the social units involved in the offense. Thus, in some societies if adultery or homicide occurs within the local lineage it may be regarded more as a sin than as a legal wrong, calling for ritual expiation rather than the application of jural sanctions. If, however, it involves members of different clans within a larger political unit, the offense may be met by a punitive raid or may have to be compounded by payment of compensation. In other words, it is not enough to regard a given sanction simply as the appropriate response to, or mode of redress for, any particular offense. Rather the sanction has to be viewed as it pertains to the set of social relationships involved in any given situation and at a particular moment in time. Failure to see the sanction in these terms can lead to misconceptions and to misguided hypotheses of an either/or character, as in attempts, for example, to correlate the importance of sorcery with the presence or absence of "superordinate justice" (Whiting 1950). Formulations of this kind pose a false dichotomy; they also miss the point that because each individual is the focus of a number of different social relationships, combining in his person a number of different roles, a variety of sanctions may attach to a single act of behavior. Thus, to take an example from our own society, a member of one of the professions who commits a serious breach of professional etiquette faces the possibility not only of prosecution in the courts, but also of being struck off the roll; he may also find that he has sacrificed his position as chairman of the local literary and philosophical society and that his wife and children have lost the good will of friends and neighbors. In other words, there is in all societies a coincidence of sanctions, some of a jural, some of a mystical or other character, all operating simultaneously toward the goal of achieving social control. Thus the problem that has to be faced is only partly the clear demarcation of one kind of sanction from another; more important is the need to classify and analyze the typical situations in which

the different kinds of sanctions are provoked or achieve dominance.

**Analytic examples.** Some progress in examining the social context of sanctions has been made by Turner (1957) in an analysis of the conflicts operating within the social system of the Ndembu of Northern Rhodesia. Among the various redressive mechanisms known to the Ndembu are informal arbitration, formal courts under the control of chiefs, and certain kinds of public ritual. Turner has suggested that jural machinery is likely to be invoked when a dispute is conducted in terms of an appeal to common norms, or, alternatively, when the dispute itself involves an appeal to different or opposed norms, but it can still be settled fairly easily because there exists "a common frame of values which organize a society's norms into a hierarchy." There are, however, other kinds of conflict, which have their source in the contradictions inherent in the social structure itself. For example, when the rules of nepotic and adelphic succession both operate within a matrilineal system, they tend to promote faction within local groups. If these conflicts are not effectively sealed off, they may lead to the disruption of the group. For disputes of this kind, usually sparked by the breach of some legal or customary norm, jural sanctions are likely to be inappropriate and ineffectual. A court of law, employing rational techniques of inquiry, hears the arguments of the various parties and on its reading of the evidence proceeds to allocate blame to one or the other of the disputants. Such a procedure cannot "relieve the quarrels so as to preserve the threatened relationships." This explains why it is that among the Ndembu and other peoples of the central African "matrilineal belt" so few cases arising out of disputes within the matrilineage come before the tribal courts. These indeed are the kinds of situations par excellence in which the sanctions associated with accusations of witchcraft and sorcery or the wrath of ancestral spirits are predominant and where group unity can be reaffirmed only by ritual means. So, too, in modern Western society the invocation of legal sanctions may be found inadequate for the handling of certain kinds of domestic disputes. A court of law, by concentrating on husband and wife as right-bearing and duty-bearing units, excludes itself from examining a close and complex relationship in its totality; the less rigid techniques of a marriage guidance council or other similar agency may have to be invoked if reconciliation of the parties is the end to be achieved.

With the development of recognized courts, a new and important phase has been reached in the



eternal task of achieving social control. Once forensic institutions have been established, little comes to escape their purview, and there are few matters that may not be the subject of judicial scrutiny. Yet, perhaps for this very reason, it is important to remember, as the last example cited itself suggests, that adjudication may cover a range of different procedures. The distinguishing criteria of a court appear to be that it is capable of enforcing the principle of jurisdiction and that it is led to the apportionment of liability by the rational assessment of the evidence and arguments adduced before it.

In many African societies, informal hearings before village elders and the detection of the causes of personal and natural misfortune by divination share certain of the characteristics of the judicial process, but clearly it would be a mistake to equate them with courts. Similarly, as Gluckman (1962) argues, it seems a mistake to speak of the institutionalized song duels of the Eskimo as "juridical instruments" even though they do serve to settle disputes and restore normal relations between estranged members of the community. It may be true that the result of the contest is a "judgment" in favor of one of the contestants, but, so far as can be gathered from the literature, the mode of arriving at that judgment has little in common with a judicial process. Here a lack of clear analytical distinctions can blur the important sociological problem of why in different societies, or in different ranges of relationship within the same society, different kinds of sanctions may be invoked for similar kinds of breach of norm.

Nevertheless, the instance of the Eskimo song encounter does direct attention to a variety of sanctions occupying a borderland between the realms of law, morals, and custom. The literature on the song contest does not make clear how points are scored in the duel, nor does it indicate what part the spectators play in this connection. On the other hand, the fact that the contestants are obliged to accept the taunts, jibes, and accusations of their opponents without taking affront suggests parallels with forms of licensed "joking" familiar in the anthropological literature. There is, for example, the system of clan joking relationships described for the Plateau Tonga of Northern Rhodesia. Among the Tonga, the clans are the only enduring units of social organization, and they are linked in a complex arrangement of joking partnerships. The Tonga explain these relationships by reference to some antagonism between the animals from which the clans derive their names or to some legend of origin. Whenever members of such paired clans

meet, they tease one another and indulge in mutual abuse. Joking of this kind occurs in informal situations, but this, like the Eskimo song contests, which are conducted for pure entertainment, has to be seen as a necessary preparation of the stage, so to speak, for those more serious occasions when culturally licensed abuse may be employed as a public sanction. The jural unit of Tonga society is the matrilineal group, but there are many situations where this group is not free to act without destroying the very principle of lineage unity on which the social structure is built. Those who waste their kin's property or commit incest provide instances of this kind. Their offenses are, in a sense, jural wrongs; but they also touch questions of ultimate social morality. For the protection of the community, the sanction of public castigation and ridicule of the offender by his clan joking partner is invoked (Colson 1953). Institutions embracing the "joking" principle are widespread; and they are not confined to tribal societies. It appears that "youth groups" may perform a similar role in some of the rural communities of present-day Spain and Wales (see Peters 1967). Such groups, which initially establish their "irresponsibility" by playing childish pranks as part of their normal behavior, are on occasion able to express the outraged moral feelings of the community by playing offensive practical jokes on those who have offended but who cannot be readily reached by straightforward legal sanctions. These instances, incidentally, illustrate another important mechanism in the process of social control. This is the principle whereby the enforcement of the sanction is thrust upon a "stranger" to the group; it is used in situations involving the unity of the group rather than the claims of aggrieved individuals. Through the medium of the clan joking partners, or the youth group, who are strangers to the adult community by virtue of their customary infantile behavior, the fiction of group cohesion can be maintained and group norms reaffirmed.

**Function of integration.** The importance of the sanction lies, then, as Radcliffe-Brown noted (1934), in its integrative function. The social system is made up of many different kinds of groupings, frequently based on different principles of social organization and serving different sets of interests. Associated with these groupings and with the different and sometimes conflicting sets of social relationships in which their members are involved, there is a wide range of social sanctions, at the back of which are the organized sanctions of the law, supported by the coercive authority of society. Even in those systems where "legal totali-

tarianism" has been carried furthest, it is clear, as Lewin (1947) has documented with reference to South Africa, that social control is not achieved exclusively through the operation of the sanctions of the law. Faced with this problem, certain jurists have been led to postulate a "general habit of obedience" to explain why, everywhere in society, laws are more often obeyed than disobeyed. A sociological approach, on the other hand, would stress that the efficacy of law rests, in the final analysis, on the support it receives from the operation of non-legal sanctions in the various domains of social life.

We can see something of the way this process operates in some analyses of the judicial process in African tribal societies. Barotse judges, for example, had behind them the full power and authority of their nation as well as of the British colonial administration. In the handling of disputes, however, the sanctions of morality and custom and, indeed, the whole normative structure of the society are channeled into the judicial process itself. Litigants and judges alike appeal to, and couch their arguments in terms of, the same body of norms. When a man is condemned, it is usually because he has failed to measure up to the standards of behavior that he himself has invoked in his statements to the court (Gluckman 1955). It is, of course, patent that there are important differences between the judicial processes in African societies and those in modern Western societies. Unfortunately the classic studies of the judicial process in Anglo-American systems have concentrated on judicial reasoning in the upper levels of the judicial hierarchy or upon exposing the nonlegal influences that play upon the mind of the judge. What seems to be required for purposes of significant comparison is the more intensive sociological investigation of courts and other similar bodies at the lower levels of the hierarchy, in the context of the communities within which they operate. Such studies might prove to be very important in illuminating the way in which social norms and sanctions are brought into relationship with the legal system.

A. L. EPSTEIN

[Directly related are the entries JUDICIAL PROCESS; LAW; POLITICAL ANTHROPOLOGY.]

#### BIBLIOGRAPHY

- BOHANNAN, PAUL 1957 *Justice and Judgment Among the Tiv*. Published for the International African Institute. Oxford Univ. Press.
- COLSON, ELIZABETH (1953) 1962 Clans and the Joking-relationship Among the Plateau Tonga of Northern Rhodesia. Pages 66-83 in Elizabeth Colson, *The Plateau Tonga of Northern Rhodesia*. Manchester Univ.

- Press. → First published in Kroeber Anthropological Society, *Papers*.
- GLUCKMAN, MAX 1955 *The Judicial Process Among the Barotse of Northern Rhodesia*. Manchester Univ. Press; Glencoe, Ill.: Free Press.
- GLUCKMAN, MAX 1962 *African Jurisprudence*. *Advancement of Science* 18:439-454.
- GLUCKMAN, MAX 1965 *Politics, Law and Ritual in Tribal Society*. Oxford: Blackwell; Chicago: Aldine.
- HOEBEL, E. ADAMSON 1954 *The Law of Primitive Man: A Study in Comparative Legal Dynamics*. Cambridge, Mass.: Harvard Univ. Press.
- LEWIN, JULIUS 1947 *Studies in African Native Law*. Cape Town: African Bookman; Philadelphia: Univ. of Pennsylvania Press.
- LLEWELLYN, KARL N.; and HOEBEL, E. A. 1941 *The Cheyenne Way: Conflict and Case Law in Primitive Jurisprudence*. Norman: Univ. of Oklahoma Press.
- MALINOWSKI, BRONISLAW (1926) 1961 *Crime and Custom in Savage Society*. London: Routledge. → A paperback edition was published in 1959 by Littlefield.
- NADEL, SIEGFRIED F. 1947 *The Nuba: An Anthropological Study of the Hill Tribes in Kordofan*. Oxford Univ. Press.
- PETERS, E. L. 1967 *Group Joking Relationships*. Unpublished manuscript.
- POSPISIL, LEOPOLD 1958 *Kapauku Papuans and Their Law*. Yale University Publications in Anthropology, No. 54. New Haven: Yale Univ.
- RADCLIFFE-BROWN, A. R. (1922) 1948 *The Andaman Islanders*. Glencoe, Ill.: Free Press.
- RADCLIFFE-BROWN, A. R. 1934 *Social Sanction*. Volume 13, pages 531-534 in *Encyclopaedia of the Social Sciences*. New York: Macmillan. → Reprinted in Radcliffe-Brown's *Structure and Function in Primitive Society*, published by Cohen and West in 1952.
- SEAGLE, WILLIAM (1941) 1946 *The History of Law*. 2d ed. New York: Tudor. → First published as *The Quest for Law*.
- TURNER, V. W. 1957 *Schism and Continuity in an African Society: A Study of Ndembu Village Life*. Manchester Univ. Press.
- WHITING, BEATRICE B. 1950 *Patute Sorcery*. Viking Fund Publications in Anthropology, No. 15. New York: Viking Fund.

#### SANCTIONS, INTERNATIONAL

The term "sanction," in its widest sense, means "any measure taken in support of a social order regulating human behavior" (Kelsen 1957, p. 101). The purpose of a sanction is to bring about a behavior considered to be in conformity with the goals and standards of a society and to prevent that behavior which is inconsistent with these goals and standards. Consequently, in its widest sense, a sanction can have the objective of suppressing and terminating a particular form of undesirable behavior and the objective of deterring or discouraging such behavior in the future. The word "sanction" originated in Roman law, where it meant a penalty imposed upon a person who violates the law. The word continues to be used in this narrow



sense, particularly in municipal law. In international relations, the term is customarily used in the wider sense.

It is sometimes asserted, particularly with reference to Austin's definition of law, that international law is not law because it lacks the sanctions of municipal law. That there is a difference of degree is no doubt true, but what is often not recognized is that municipal law, like international law, is basically dependent upon custom and consent, while international law has sanctions comparable with those of municipal law, though admittedly weaker. The organization of sanctions in support of an international order faces a special difficulty not encountered within the state to anything like the same degree. This is the difficulty of enforcing standards against organized groups which, as states, make exclusive claims on governmental authority and on the loyalty and support of their citizens. Even within the state, constituted authorities experience difficulties in enforcing their decisions against organized groups such as trade unions and component political units, as witness the difficulties experienced by the organs of the federal government of the United States in enforcing their decisions against the states.

### Experience

Customary international law has in the past admitted the right of states to take certain measures of self-help in support of their legal claims if it was not possible to achieve satisfaction by negotiation or third-party settlement. These measures included withdrawal of diplomatic representatives, acts of reprisal such as embargo and pacific blockade, display of force, landing of armed forces for purposes of protection of persons and property, and resort to war. Under agreements concluded to meet particular situations, special guarantees were accorded and the right of intervention admitted, as, for example, in the collective guarantee of Belgian neutrality (1839) and the United States guarantee of Cuban independence (1901).

Under the Covenant of the League of Nations, the concept that peace should be maintained by the combined forces of peace-loving states was substituted for the principles of balance of power and self-help. "Resort to war" under defined conditions was prohibited. Provision was made under article 10 for joint guarantee of member states against external aggression and in article 16 for sweeping economic and financial sanctions to be applied "immediately" by every member against any state resorting to war in violation of its obligations under the Covenant. The Covenant provisions were seri-

ously weakened by resolutions adopted by the Assembly in 1921 that were intended to take account of the failure of the League to achieve universality. Military sanctions under the Covenant were optional on recommendation of the Council.

No serious consideration was given to the application of the sanctions of article 16 until the mid-1930s. The Japanese invasion of Manchuria in 1931 was not treated as a resort to war under the Covenant. The Assembly did, however, recommend to members that they withhold recognition of the fruits of Japan's military action, in line with the announced policy of the United States. This constituted a mild form of sanction, which had, however, no observable practical effect. In 1935, however, Italy's military invasion of Ethiopia was recognized by all except a small minority of members as an illegal resort to war, and limited economic and financial sanctions were applied. These were not adequate, however, to achieve the declared objective of frustrating Italy's military action, and leading members of the League were unwilling, for political reasons, to agree to measures that might have been effective.

The Charter of the United Nations adopts a different approach to the problem of sanctions. Like the Covenant, its objective is limited to the maintenance of peace; the Charter prohibits any threat or use of force against the territorial integrity or political independence of any state "or in any other manner inconsistent with the Purposes of the United Nations" (art. 2, paragraph 4). The sanctions envisaged, however, are not automatic, being left to the discretion of the Security Council. In cases of threats to the peace, breaches of the peace, or acts of aggression, it may take decisions requiring members to apply collective measures, including the severance of diplomatic relations, economic and financial nonintercourse, and "action by air, sea or land forces" (art. 42). But the taking of military measures is made dependent on the conclusion of agreements by which members undertake to place forces and facilities at the disposal of the Security Council (art. 43). In addition, a member against whom preventive or enforcement action has been taken may be suspended from "the exercise of the rights and privileges of membership" by the General Assembly upon the recommendation of the Security Council (art. 5), and a member who has "persistently violated the Principles" contained in the Charter may be expelled in like manner (art. 6). The Charter does not confer upon the General Assembly any explicit powers with respect to the imposition of sanctions. It does give to the Assembly the general authority to discuss and make

recommendations with respect to any matter within the scope of the Charter (art. 10).

In United Nations practice, the Security Council has largely been prevented from using its powers to apply sanctions—its enforcement powers—by the inability of the permanent members to agree. In the early stages of its handling of the Palestine question (1948) it specifically invoked these powers and threatened to use them. In the Korean case, the Council, in the absence of the Soviet representative, recommended assistance to the Republic of Korea and, under this authorization, members took measures of the kind described in articles 41 and 42. More recently, in dealing with policies of racial discrimination by South Africa, the Security Council, by its resolution of December 4, 1963, called upon members to refrain from shipping arms and munitions to that country so long as it practiced this discrimination, which it found was "seriously disturbing international peace and security."

The General Assembly, in the exercise of its broad powers under the Charter, has on numerous occasions recommended to members the adoption of various measures of a coercive nature to bring the conduct of particular states into line with standards set by the Charter, international law, and United Nations resolutions. Examples include such recommended measures as the withdrawal of heads of missions to Spain and its exclusion from UN conferences in order to eliminate the Franco regime (1946); an embargo on the shipment of arms and strategic materials in general to North Korea and Communist China to suppress their aggressive action (1951); and an embargo on export of arms and petroleum products to South Africa to achieve the termination of policies of racial discrimination (1963). The measures recommended by the General Assembly have had a range of purposes extending beyond the maintenance of international peace and security, to the implementation of obligations and policies in respect to self-determination and human rights.

Provisions for the application of international sanctions are not limited to the constitutions of global international organizations. The Inter-American Treaty of Mutual Assistance, for example, provides that in case of aggression by an American state, the Organ of Consultation may decide that the other signatory states apply sanctions, including political, economic, financial, and military measures. Decisions may be taken by a vote of two-thirds and are binding, except that no member can be required to use military force against its will. In August 1960, the decision was taken to sever diplomatic relations and partially interrupt

economic relations with the Dominican Republic because of its aggression against Venezuela.

### Evaluation

Experience with organized international sanctions in support of international order has been limited and does not give much encouragement to those who believe in the efficacy of such arrangements. The League system of sanctions was a system of collective security which rested "upon the proposition that war can be prevented by the deterrent effect of overwhelming power upon states which are too rational to invite certain defeat" (Claude 1964, p. 228). It also assumed that an ethical and legal difference existed between military force used by a state to advance its national interests and the use of force by members of an international organization to restrain a state after it has been adjudged guilty of illegal use of force. This experiment was handicapped from the beginning by the League's lack of universality. The limited economic and financial sanctions imposed against Italy showed that international sanctions were technically feasible. The necessary machinery could be organized and decisions could be taken to permit sanctions to become operative promptly and with substantial impact on the transgressor. Their failure was due to political difficulties—the conflicting interests and purposes of the sanctioning powers and the unwillingness of some states to risk war to make sanctions effective.

United Nations experience has not been encouraging either. Sanctions against Franco Spain were applied half-heartedly and were soon terminated. In the Korean case, the measures taken to assist the Republic of Korea and to defeat the aggressive actions of North Korea and Communist China were essentially United States measures dictated by United States policy considerations. The United Nations served as a useful framework to provide legitimacy and as a moderating influence. Measures recommended against South Africa to achieve fuller respect by that member for basic human rights have had little observable effect.

What conclusions are to be drawn from this and other experience regarding the role of international sanctions in the implementation of international order? Statesmen and scholars have arrived at a variety of conclusions.

One conclusion which appears to be widely held is that the theory of international sanctions, particularly as it involves the use of military force, is fallacious. According to this view, no effective pressure can be applied against a state except by another state, and consequently, an international



order that depends on that form of pressure is a permanent source of conflict instead of a medium of order. As early as 1787 Hamilton wrote in *The Federalist*: "In an association where the general authority is confined to the collective bodies of the communities that compose it, every breach of the laws must involve a state of war; and military execution must become the only instrument of civil obedience" (No. 15).

Another conclusion that has wide support is that forces of coercion which are effective when applied to individuals are not effective, at least not to the same degree, when applied to states. Even those forms which might be expected to be particularly effective, such as embargoes and blockades, may at least temporarily have the opposite effect. In the case of the economic sanctions against Italy, among the immediate effects was a strengthening of popular support for the regime.

A view which is widely held by those who stress national interest as the motivating force in international relations is that international sanctions are unrealistic because they require a willingness to accept general commitments for situations which cannot be foreseen in detail and to carry them out under circumstances where the immediate national interest appears to dictate a different course. Thus, in the case of League sanctions against Italy, France was a reluctant participant because of her primary interest in having Italian support against Germany. Many members of the United Nations were reluctant to press sanctions against Communist China in 1950 and 1951 because of their greater concern at the time with the security of western Europe.

A more optimistic and positive view regarding sanctions is that while experience to date may not have been encouraging, a system of organized sanctions in support of law and order is the only alternative to anarchy. Scholars and writers who take this view—Quincy Wright and Myres McDougal are examples—argue that our experience with sanctions as applied to organized groups is not so discouraging as is often contended, that sanctions have been effective in deterring undesirable conduct by states, and that with a fuller appreciation of the extent of the coincidence of particular and general interests in public order, sanctions can be expected to become more effective.

### Problems and prospects

It is obvious that the organization and application of international sanctions in support of international order present difficulties that do not exist to the same degree in the case of sanctions in sup-

port of domestic order. With respect to international legal claims, states generally continue to rely upon good faith and reciprocity of interest in compliance, and upon retaliatory measures when these fail. It is primarily for the maintenance of peace and, to an increasing extent of late, for the implementation of certain human rights that organized sanctions are seriously considered. Much study has been given to the organization and application of collective sanctions to these ends, both by private scholars and by official groups.

A group of members of the Royal Institute of International Affairs made a study of international sanctions in the light of the League experience of 1935–1936 (1938). During the period when the United Nations was engaged in supporting collective measures in Korea, the General Assembly, by its Uniting for Peace Resolution of November 3, 1950, established the Collective Measures Committee to study methods which might be used to strengthen international peace and security. In its report, published in 1951, it made detailed recommendations for the organization and application of collective measures (United Nations . . . 1951).

The scope of the prohibition of the use or threat of force under the Charter, combined with the lack of any effective provision for the enforcement of legal rights, has led some scholars to take the view that unilateral use of force to secure respect for legal rights is permissible as a sanction consistent with "the Purposes of the United Nations" (Stone 1958, pp. 94–95). Others take the view that the Charter does not permit the individual use of force, without United Nations authorization, except as it can be justified as an exercise of the right of self-defense in case of an armed attack.

There is an extensive scholarly literature dealing particularly with the theory and legal aspects of international sanctions, but research and writing have thus far neglected the more practical aspects of the problem. While scholars have explored the legal, organizational, and administrative problems involved in the establishment and operation of a system of sanctions, and furthermore have analyzed the influence of the international political environment on a system of sanctions, relatively little attention has been given to the effects of sanctions upon the delinquent states as well as upon those states joining in their application. It is too readily assumed that the recalcitrant state can be coerced by material pressures into conforming to a desired mode of conduct or into ceasing an undesirable mode of conduct. We need to know more about the effect of sanctions upon the people of the state that is the object of sanctions, and the

processes by which that state, once its resistance is overcome, can be restored to good standing in the international community.

Furthermore, while the problem of the unequal effects of sanctions upon those states applying them or expected to apply them is recognized, more study needs to be made of the extent of these inequalities and of procedures by which burdens will be more equally shared. It is clear that reluctance to participate in the application of sanctions has in large measure been due to unequal burdens and risks that states have been asked to assume.

LELAND M. GOODRICH

[See also ECONOMIC WARFARE; INTERNATIONAL CONFLICT RESOLUTION; INTERNATIONAL LAW. Other relevant material may be found under INTERNATIONAL ORGANIZATION; INTERNATIONAL POLITICS; WAR.]

#### BIBLIOGRAPHY

- BRIERLY, J. L. 1931 Sanctions. Grotius Society, Transactions 17:67-82.
- CLAUDE, INIS L. 1964 *Swords into Plowshares: The Problems and Progress of International Organization*. 3d ed. New York: Random House. → The first edition was published in 1956.
- GOODRICH, LELAND M.; and SIMONS, ANNE P. 1955 *The United Nations and the Maintenance of International Peace and Security*. Washington: Brookings Institution.
- KELSEN, HANS 1957 *Collective Security Under International Law*. U.S. Naval War College, International Law Studies, Vol. 49. Washington: Government Printing Office.
- KUNZ, JOSEF L. 1960 Sanctions in International Law. *American Journal of International Law* 54:324-347.
- MCDUGAL, MYRES S.; and FELICIANO, FLORENTINO P. 1961 *Law and Minimum World Public Order: The Legal Regulation of International Coercion*. New Haven: Yale Univ. Press.
- NIEMEYER, GERHART 1941 *Law Without Force: The Function of Politics in International Law*. Princeton Univ. Press.
- ROYAL INSTITUTE OF INTERNATIONAL AFFAIRS 1938 *International Sanctions: A Report by a Group of Members of the Royal Institute of International Affairs*. Oxford Univ. Press.
- STONE, JULIUS (1954) 1959 *Legal Controls of International Conflict: A Treatise on the Dynamics of Disputes and War Law*. 2d impression, rev., with supplement 1953-1958. New York: Holt.
- STONE, JULIUS 1958 *Aggression and World Order: A Critique of United Nations Theories of Aggression*. Berkeley: Univ. of California Press.
- TAUBENFELD, RITA FALK; and TAUBENFELD, HOWARD J. 1964 The "Economic Weapon": The League and The United Nations. *American Society of International Law, Proceedings* [1964]:183-205.
- UNITED NATIONS, GENERAL ASSEMBLY 1951 Report of the Collective Measures Committee. United Nations, General Assembly, Official Records Sixth Session, Supplement no. 13.

## SAPIR, EDWARD

- I. CONTRIBUTIONS TO CULTURAL ANTHROPOLOGY David G. Mandelbaum  
II. CONTRIBUTIONS TO LINGUISTICS Z. S. Harris

#### I

##### CONTRIBUTIONS TO CULTURAL ANTHROPOLOGY

Edward Sapir (1884-1939) was one of the founders of modern linguistic anthropology, a main contributor to the development of formal descriptive linguistics, a leading figure in cultural anthropology, and a chief stimulator of studies in the relations between personality and culture. His ideas continue to excite lively interest; his writings are widely read; his one book, *Language*, and a volume of his papers have continuously been reprinted. The range of his ideas is great. They are founded on an extraordinary breadth of knowledge, which Sapir used with incisive power of thought and subtlety of analysis. He wrote poetry, literary essays, and music as well as scholarly works; all his writings are characterized by an unusual felicity of style.

Sapir was born in Pomerania (Germany) and was five years old when his family emigrated to the United States. His intellectual gifts were recognized while he was a schoolboy in New York City. He was awarded scholarships to the Horace Mann School and then a Pulitzer fellowship to Columbia College, from which he was graduated in 1904. He went on to do a year of graduate work in Germanics at Columbia, and at that time his interest in language brought him into contact with Franz Boas. He quickly learned from Boas the great potentialities of the anthropological study of language, and from then on his principal work was in linguistic anthropology. [See the biography of Boas.]

In 1910 Sapir went to Ottawa as chief of the division of anthropology in the Geological Survey of Canada. His 15 years in that post gave him fine opportunities for field research and writing, but it was also a period of relative isolation from the main centers of anthropological work. He welcomed the invitation to join the department of anthropology at the University of Chicago, all the more because his wife had died after a long illness. He came to Chicago with his three children in 1925 and remained at the university for six years; his work and his influence flourished there. He was married again, to Jean McClenaghan, and they had two children. In 1931 he accepted a Sterling professorship at Yale and founded the department of anthropology there. His years at Yale were active and productive until he suffered a series of heart



attacks in 1937, which ended his life two years later.

Sapir's primary work was always in the study of language (his achievements in that field are discussed in the second section of this biography). His writings on other subjects, such as culture, society, and personality, were often done in response to a request for an article on a particular topic or as offshoots of his linguistic studies. Yet a good many of them have been major contributions which still command attention. These writings are of three main kinds—ethnological studies of American Indians, essays on general concepts, and papers on personality in culture and society.

**Ethnological studies.** Sapir's ethnological contributions appeared in steady succession from his earliest writings to some that were published posthumously. Many were brief studies of the kind that provide the necessary groundwork for more general formulations, as the account of Takelma religion, or of a Nootka puberty ceremony, or of Sarsi pottery. In some, a large body of data is summarized and underlying principles demonstrated. For example, in "The Social Organization of the West Coast Tribes" (1915) Sapir showed how the idea of social ranking pervaded many aspects of these societies—how the emphases on inherited privileges and on ritual insignia were linked to that prevailing idea. When Robert Lowie wrote an appraisal of Sapir's work long after Sapir's death, he noted that Sapir was one of the very best of ethnographers—able to immerse himself in the phenomena under scrutiny, plumb them to the depths, and present them deftly and incisively (*Letters From Edward Sapir to Robert H. Lowie*, p. 10).

Sapir's most substantial contribution to ethnological theory is his monograph of 1916, "Time Perspective in Aboriginal American Culture: A Study in Method." A main objective of field research among American Indians at that time was understanding the actual processes of culture change as against the presumed universal processes, which had been postulated by nineteenth-century evolutionists. In order to understand how American Indian cultures had actually changed, anthropologists had to reconstruct cultural histories. In this monograph Sapir surveyed the methods useful in such reconstructions; he particularly illuminated the use of inferential evidence, such as that from the study of cultural associations, geographic distributions, cultural strata, and from grammatical and lexical analyses. He presented the logical basis for evaluating different kinds of inferential evidence. The arguments are

closely reasoned; supporting evidence is compactly presented; broad concepts are tersely stated.

It is a work that requires close study, and generations of graduate students studied it closely because it was for many years one of the major theoretical pieces on American Indian anthropology. Some of these students found that a single paragraph or even a sentence could open fruitful leads for their own research. In recent years, the shift of anthropological attention to peoples in other parts of the world and the development of interest in other theoretical approaches have lessened the importance of this work, but it held a leading place for some forty years.

**General concepts.** Sapir's less technical, more general essays have had a wider audience; they have a way of including some ideas which appear, decades later, to be at the frontiers of thought. Thus, in "Culture, Genuine and Spurious" (1924), Sapir defined the anthropologist's characteristically objective view of a people's way of life and then went on to distinguish genuine and worthy cultures from spurious and morally shoddy ones. He valued those cultures that encourage harmonious and self-satisfying lives for the people who live by them, whatever the level of technical achievement and efficiency. He decried those life styles, as of some modern industrial and service workers, in which a person finds little relationship or reciprocal gratification among his various roles and pursuits. This thesis, although strange to anthropology, was true to the intellectual temper of the time. Sapir's special touch is found not only in the evidence he selected and the telling way in which he presented it, but even more in the intriguing ideas he mentioned, as it were, in passing. Thus, he commented on the nexus of self and society, on the place of art in a genuine culture, on the cultural pressures generated by political forces. Of this last matter he wrote, "Our national-political units are too small for peace, too large for safety" ([1924] 1949, p. 330). There is a good deal in this essay that belongs to the 1920s, but there is also a good deal that holds interest still.

Other essays clarify fundamental concepts and categories. In "The Meaning of Religion" (1928) Sapir took a subject which, in other hands, is often hackneyed; in his, the subject is treated freshly and lucidly, with proper anthropological objectivity and yet with sympathy. Here also he threw off a number of productive ideas, pointing out, for example, that a basic distinction among religions is exemplified by the contrast between Plains and Pueblo Indian practices. The two types tend to differ "according to whether they find the last court of appeal in mat-

ters religious, in the social act, or in the private emotional experience" ([1928] 1949, p. 350).

The article "Anthropology and Sociology" (1927a) appeared as a chapter in a book on the social sciences and their interrelations. In it Sapir was not at all concerned with separating the two domains or negotiating boundaries between them. He concentrated instead on the ideas that anthropology can contribute toward an understanding of people and behavior. He began by noting that a main error of earlier theorists lay in taking primitive people to be archaic, unchanging prototypes. Sapir expressed the view that has now become accepted, that "we are to be at least as much interested in the many points of accord between primitive and sophisticated types of social organization as in their sensational differences" ([1927a] 1949, p. 336).

These similarities, he continued, are not readily apparent from reconstructions of culture histories of American Indian tribes. It is rather from the "stupendous facts" about common human trends that sociologists will in the long run have most to learn from anthropology. The separate historical explanations are little more than a necessary clearing of the ground for a social interpretation; they are not to be mistaken for the interpretation itself. In his "Time Perspective" (1916) Sapir had provided a major theoretical foundation for the work of historical reconstruction, which then preoccupied most American ethnologists. In the article of 1927 he recognized the work he had done in 1916 as only a preliminary step toward the anthropologist's principal tasks.

Another development of Sapir's thought appears in several of the articles he wrote for the *Encyclopaedia of the Social Sciences*: an emphasis on the relation between individual and society, between personality and culture. This emphasis arose partly because of Sapir's humanist feeling that the proper study of man must include real men and partly because of his realization that formal, non-personalized formulations about people can be very misleading unless they are coupled with insights into the real experience and reactions of actual persons. In the article "Group" (1932a), for example, Sapir sketched the essential definition and dimensions of the concept in his clear, succinct way. He outlined three types of classification based on objective observation and three based on subjective engagement of the participant. He noted the bases for group formation and changes, then posed the fundamental question of the bonds that hold together the members of a group. The psychological basis of the group, he observed, must rest on the psychology of specific social relations. Loyal-

ties to an entire group do not mean that direct relationships between individual and individual have been completely transcended, but rather that the attributes of interpersonal experience have been transferred to the group as a whole. Thus the psychological realities of group participation can best be understood through studies of the actual kinds of understanding and expectation that grow up between two or more human beings when they are brought into significant contact.

**Culture and personality.** Sapir's interest in the interplay of culture and personality dated from the beginning of his anthropological studies. "I remember being rather shocked than pleased," he wrote in a paper of 1938, "when in my student days I came across such statements in J. O. Dorsey's 'Omaha Sociology' as 'Two Crows denies this'" ([1938] 1949, p. 569). The paper is devoted to an assessment of Two Crows' denial and of what the anthropologist should make of it. Sapir found conventional ethnographic statements inadequate because they were generalized formulas which said little or nothing about variation, conflict, denial, or emotion. Even as a student Sapir recognized that such generalized formulations might be necessary but could not be sufficient for a thorough understanding of behavior. So Dorsey was ahead of his age in recognizing that he was not dealing with an abstracted society or a set of specimens but with a number of human beings who could give themselves the privilege of differing from each other on matters that were, in the conventional anthropological definition, part of the official tribal culture.

Sapir was moved to point out the inadequacy of rigidly depersonalized studies in his response of 1917 to A. L. Kroeber's formulation of the concept of the superorganic. In "Do We Need a Superorganic?" he granted that few individuals, as individuals, make a marked impress on social movements, and yet it is always the individual who thinks and acts and dreams and revolts. The difference between social and individual data in behavior studies results from the observer's choice of what to select for systematic notice. Hence it is not valid to draw an abidingly sharp line between the two types of data, since "individual" reactions constantly spill over and give color to "social" reactions (1917, p. 442).

This insistence on the significance of the person appears recurrently in Sapir's writings on culture and society. His paper "Speech as a Personality Trait" (1927b) is one instance of this theme; other expressions of it came in a series of papers written in the last decade of his life.

In a paper of 1932 he restated the idea that



the true locus of culture is in the interactions of specific individuals and in the meanings which the participants abstract from these interactions ([1932b] 1949, p. 515). And in an article of 1938 he proposed a research procedure that has only lately been appreciated. Instead of arguing from a supposed objectivity of culture to the problem of individual variation, Sapir recommended that for certain kinds of analyses we should operate as though we knew nothing about culture but were interested in ascertaining what a set of human beings who are accustomed to live with each other actually think and do in their day-to-day relationships ([1938] 1949, p. 574). This simple discovery procedure, as it turned out, involves complex questions as to when and how it is to be used, but it can also lead to considerable advances in anthropological understanding.

As for the interpretation of a particular sequence of behavior as being either social or personal, Sapir's paper of 1934 on this matter shows that the duality of reference is an inevitable part of human experience as well as of the social scientist's method. When one is confronted with familiar circumstances and familiar people, interest is likely to be focused on the individual. With unfamiliar people or situations, in socially distant or casual relations, one's interest is more likely to center on cultural or social observations. Each type of interest "is necessary for the psychic preservation of the individual in an environment which experience makes increasingly complex and unassimilable on its own simple terms" ([1934] 1949, p. 591). Every person must use both frames of reference in his own conduct. In the study of man the anthropologist must also use both, but with explicit awareness of what he is doing, and he must be able to deploy each in ways best suited for analysis and explanation.

Sociology and anthropology, Sapir continued, are rooted in the social kind of reference, a kind which perhaps is characteristic of the child's earliest experience with the seemingly irresistible authority and knowledge of his parents. Psychiatry and social psychology proceed more from the personal kind of reference and from the necessity felt to assert oneself significantly. All these are "preliminary disciplines" and provide an invaluable first ordering of research.

But sooner or later their obscure opposition of spirit must be transcended for an objectivity which is not merely formal and non-evaluative but which boldly essays to bring every cultural pattern back to the living context from which it has been abstracted in the

first place and, in parallel fashion, to bring every fact of personality formation back to its social matrix. ([1934] 1949, p. 592)

For such purposes, the descriptions and classifications of abstracted culture patterns are no longer very relevant, useful as they are to clear the observational ground. The more anthropologists become concerned with interactional meanings and values rather than with segregated traits, Sapir commented, the more a culture seems to take on the characteristics of a personality organization, and the growth of culture is understood in the spirit of the development of a personality. Thus, such ethnological questions as that of clan membership may well become subsidiary to inquiries as to whether a father is in the habit of acting toward his son as an indulgent guide or as a disciplinarian. One result of taking this point of view is that the bizarre or exotic in alien cultures tends to be minimized, while the broad human base on which all culture has developed is revealed more clearly (*ibid.*, pp. 594-595). Anthropologists who appreciate that a culture is not given in a neat package to the child, but is gropingly discovered and assimilated by him, and who can see that a people change their culture with similar groping assimilation, will join with psychologists in phrasing more fruitful questions about culture and behavior ([1932b] 1949, p. 521; [1934] 1949, p. 597).

**Influence.** Sapir taught students of anthropology for only about a dozen years, but his influence on several generations of anthropologists and linguists has been considerable. Hearing him in lectures or discussions could be an exhilarating experience. He was not at all dramatic; his talk was low-keyed and anything but hortatory. But he could give his students glimpses of intellectual vistas the like of which they had neither seen before nor would soon forget.

He came to differ with his teacher in anthropology, Franz Boas, on a number of fundamental matters, but Boas always maintained a high respect for Sapir's work. Colleagues and friends of his own generation showed how much they owed to his work by the formal honors they gave him during his life and by the appreciations they wrote after his death at age 56. Parts of his correspondence with two of his friends, Robert H. Lowie and Ruth Benedict, have been published (Lowie & Sapir 1965; Benedict 1959, pp. 45-54, 158-197). These letters between friends contain private banter, trade gossip, and transient comment. But they also tell something of the day-to-day context of Sapir's

life and work—for example, at one period in Chicago he was teaching as much as fourteen hours a week—and they reflect something of both the warm charm and the occasional friction in his personal relations. The letters to Lowie show Sapir in the course of developing ethnological and linguistic ideas, sometimes surely and swiftly, sometimes haltingly and piecemeal. Those to Ruth Benedict show his sharp insight into the world of poets and poetry, a world seemingly very far removed from the technical universe which he discussed with Lowie but actually interwoven in his personality and his work.

In her account of Ruth Benedict's poetry, Margaret Mead has noted that "it was in the vivid, voluminous correspondence with Edward Sapir that her own poetic interest and capacity matured." And in another passage Margaret Mead mentions that among those who benefited most from Sapir's speculations about personality and culture were Ruth Benedict, John Dollard, and herself (Benedict 1959, pp. 90, 209). The memorial volume for him written by some of his students (*Language, Culture, and Personality* 1941) was reprinted in 1960. In the generation of the students of his students, there is not only quite lively interest in what Sapir had to say, but also a good many who are urging their students to read Sapir.

Sapir's death cut short a number of scholarly enterprises, among them the analysis of a large collection of American Indian ethnological data, extensive explorations in Indo-European linguistics, preliminary formulations of linguistic relationships on the largest scale, and extensive ethnological and linguistic notes in Semitics. Toward the end of his life he turned increasingly to Semitic studies. His father had been a cantor, and he had been raised in the strict tradition of Orthodox Judaism. Although he soon discarded what he felt to be the intolerable restrictions of orthodoxy, he retained his delight in the scholarship and spirit of Judaism. His strong commitment to human rights, opposing all invidious discrimination and oppression, appears in his writing, especially in articles done for general audiences. As the grave world events of the 1930s developed, he participated actively in the defense of these rights.

The weight of his scholarly influence is in linguistic anthropology, where the weight of his published contributions also lies. In American Indian ethnology his ideas on method and leads for further research have had notable effect. In the field of culture-personality studies he remains one of the principal pathfinders. For anthropology in gen-

eral he continues to be a heartening example of how much can be accomplished when a superb mind plays over a wide and richly yielding field.

DAVID G. MANDELBAUM

[See also ANTHROPOLOGY; COMMUNICATION; CULTURE, article on THE CONCEPT OF CULTURE; INDIANS, NORTH AMERICAN.]

#### BIBLIOGRAPHY

The bibliography for this article is combined with the bibliography of the article that follows.

## II

### CONTRIBUTIONS TO LINGUISTICS

Sapir was, with Leonard Bloomfield, a founder of formal descriptive linguistics and the distributional method that characterizes it. He developed phonemic theory, the analysis of the sounds of a language according to the pattern of their distribution—that is, the sequence in which sounds occur and are systematized and perceived by the speakers of the language. Phonemic analysis thus recognizes language as behavior resulting from a process of selection.

This method of analysis classifies each part of an utterance (sound, morpheme, etc.) on the basis of its environment rather than on the basis of its meaning, phonetic properties, and so forth. The distributional method is important not only because it is self-contained but also because its universe and its terms of reference are so explicit that one can discover all sorts of hidden relations that are observable only as second-order disturbances in the overt distributional system: masked and combined phonemes, neutralized phonemic differences, zeroed morphemes, and the like. To be sure, Sapir did not always formulate aberrant phenomena in distributional terms; he sometimes used formulations drawn from biological processes. While this reduced the incisiveness and generality of the distributional theory he was developing, it did indicate some of the other systemic factors (factors of equilibrium, of dynamics, etc.) that may be required to understand the sources of aberrant phenomena or of the place these phenomena occupy in the descriptive system. These additional factors may even aid ultimately in understanding the source of the system itself. Thus Sapir might say that if, in a given language, a two-vowel sequence never occurs inside a morpheme and only rarely occurs when two morphemes adjoin each other in a word, the language will develop a "protective mechanism" to avoid the two-vowel sequence. Examples of such



mechanisms are the glottal stop and the French liaison *t*.

Distributional methods came to be widely used in the formulation of linguistic process. Phonemic phenomena were regarded as processes of change (for example, "suppletion," or the replacement of stems in certain word groups, e.g., in *am*, *is*, *be*) that are structural and thus related to the environment in which they occur. Sapir was also one of the originators of the concept of the morphophoneme. The morphophoneme classifies together all those phonemes that replace each other in the paradigm—that is, the different phonemic shapes that a word or morpheme has as it occurs in different morphemic environments. This concept was also based partly on the speaker's perception of phonemic replacement (Sapir 1933; Sapir & Swadesh 1939).

Sapir also did much work in historical and comparative linguistics, tracing the genetic relationships among languages. His interest in pattern and process led him to speak of configurational pressure as a source of change. By configurational pressure he meant that linguistic change (phonemic and morphemic) is influenced by relevant patterns inherent in the language itself. He suggested that configurational pressure is a factor in what he called drift, the process that occurs in a language whose speakers have separated and can no longer imitate ("borrow from") each other, but whose linguistic habits continue in some respects to change in the same way in the separated communities (e.g., Old Norse in Denmark and Iceland). Although this suggestion has not received sufficient study, it does accord with systematizations of linguistic change.

In historical linguistics Sapir is particularly noted for a series of intricate analyses of the earliest stages of Indo-European. He based his work partly on the difficult data from Tocharian and partly on the laryngeal hypothesis that reconstructed various laryngeal consonants lost in early Indo-European with consequent changes in neighboring vowels (1937a; 1937b; 1939). Related to this work were his investigations of the similarities and connections of Indo-European to Semitic and other Mediterranean languages. Sapir thought that Indo-European and Semitic might have descended from the same parent stock. Although it was impossible to establish this hypothesis by reconstructions using the accepted comparative method, he could point to many unexpected connections and to a great over-all structural similarity in the earliest stages of both Indo-European and Semitic.

Sapir worked with examples of structural simi-

larity in his famous attempt to group the extremely numerous and apparently unrelated families of North American Indian languages into six superfamilies (1929). Most of Sapir's linguistic work (like his ethnological studies) was with American Indians, and his mastery of the data and his analyses of the Indian languages were unequalled. He worked particularly in Navajo, Yana, Nootka, Tlingit, Sarsi, Kutchin, Chinook, Ingalik, Hupa, and Southern Paiute.

Z. S. HARRIS

[See also LANGUAGE; LINGUISTICS, article on THE FIELD; and the biographies of BLOOMFIELD; SAUSURE; WHORF.]

#### WORKS BY SAPIR

- (1915) 1949 The Social Organization of the West Coast Tribes. Pages 468–487 in *Selected Writings* . . . . Edited by David G. Mandelbaum. Berkeley: Univ. of California Press.
- (1916) 1949 Time Perspective in Aboriginal American Culture: A Study in Method. Pages 389–462 in *Selected Writings* . . . . Edited by David G. Mandelbaum. Berkeley: Univ. of California Press.
- 1917 Do We Need a Superorganic? *American Anthropologist* New Series 19:441–447.
- 1921 *Language: An Introduction to the Study of Speech*. New York: Harcourt.
- (1924) 1949 Culture, Genuine and Spurious. Pages 308–331 in *Selected Writings* . . . . Edited by David G. Mandelbaum. Berkeley: Univ. of California Press.
- (1927a) 1949 Anthropology and Sociology. Pages 332–345 in *Selected Writings* . . . . Edited by David G. Mandelbaum. Berkeley: Univ. of California Press.
- (1927b) 1949 Speech as a Personality Trait. Pages 533–543 in *Selected Writings* . . . . Edited by David G. Mandelbaum. Berkeley: Univ. of California Press.
- (1928) 1949 The Meaning of Religion. Pages 346–356 in *Selected Writings* . . . . Edited by David G. Mandelbaum. Berkeley: Univ. of California Press.
- (1929) 1949 Central and North American Languages. Pages 169–178 in *Selected Writings* . . . . Edited by David G. Mandelbaum. Berkeley: Univ. of California Press.
- (1932a) 1949 Group. Pages 357–364 in *Selected Writings* . . . . Edited by David G. Mandelbaum. Berkeley: Univ. of California Press. → First published in Volume 7 of the *Encyclopaedia of the Social Sciences*. Sapir also wrote the *Encyclopaedia* articles "Communication," "Custom," "Dialect," "Fashion," "Language," "Personality," and "Symbolism."
- (1932b) 1949 Cultural Anthropology and Psychiatry. Pages 509–521 in *Selected Writings* . . . . Edited by David G. Mandelbaum. Berkeley: Univ. of California Press.
- 1933 La réalité psychologique des phonèmes. *Journal de psychologie normale et pathologique* 30:247–265.
- (1934) 1949 The Emergence of the Concept of Personality in a Study of Cultures. Pages 590–597 in *Selected Writings* . . . . Edited by David G. Mandelbaum. Berkeley: Univ. of California Press.
- (1937a) 1949 Hebrew "Helmet": A Loanword and Its Bearing on Indo-European Phonology. Pages 285–288

- in *Selected Writings* . . . . Edited by David G. Mandelbaum. Berkeley: Univ. of California Press.
- (1937b) 1949 [A Book Review of] *The Ras Shamra Mythological Texts*, by James A. Montgomery and Zellig Harris. Pages 289–293 in *Selected Writings* . . . . Edited by David G. Mandelbaum. Berkeley: Univ. of California Press.
- (1938) 1949 Why Cultural Anthropology Needs the Psychiatrist. Pages 569–577 in *Selected Writings* . . . . Edited by David G. Mandelbaum. Berkeley: Univ. of California Press.
- (1939) 1949 From Sapir's Desk: Indo-European Prevo-calic s in Macedonian and The Indo-European Words for "Tear." Pages 294–302 in *Selected Writings* . . . . Edited by David G. Mandelbaum. Berkeley: Univ. of California Press.
- 1939 SAPIR, EDWARD; and SWADESH, MORRIS *Nootka Texts: Tales and Ethnological Narratives With Gram-matical Notes and Lexical Materials*. Philadelphia: Linguistic Society of America.
- Letters From Edward Sapir to Robert H. Lowie*. With an introduction and notes by Robert H. Lowie. Privately published, 1965.
- Selected Writings of Edward Sapir in Language, Culture, and Personality*. Edited by David G. Mandelbaum. Berkeley: University of California Press, 1949. → In-cludes a chronologically arranged bibliography of Sapir's writings.

## SUPPLEMENTARY BIBLIOGRAPHY

- BENEDICT, RUTH 1959 *An Anthropologist at Work: Writings of Ruth Benedict*, by Margaret Mead. Boston: Houghton Mifflin. → A selection of Benedict's writings, with extensive commentary by Margaret Mead.
- [Bibliography: Edward Sapir.] 1938 *Psychiatry* 1:154–157.
- HARRIS, Z. S. 1951 [A Book Review of] *Selected Writings of Edward Sapir in Language, Culture, and Per-sonality*. *Language* 27:288–333.
- Language, Culture, and Personality: Essays in Memory of Edward Sapir*. Edited by Leslie A. Spier, A. Irving Hallowell, and Stanley S. Newman. (1941) 1960 Salt Lake City: Univ. of Utah Press.
- PRESTON, RICHARD J. 1966 *Edward Sapir's Anthropol-ogy: Style, Structure, and Method*. *American Anthro-pologist New Series* 68:1105–1128.

## SARTON, GEORGE

George Sarton, historian of science, was born in Ghent, Belgium, in 1884 and died in Cambridge, Massachusetts, in 1956. He married Eleanor Mabel Elwes, an English artist, in 1911; their daughter, May, who became a poet and novelist, was born in 1912.

As a youth Sarton wrote romantic poems and essays. At the University of Ghent he studied the natural sciences and mathematics, winning a medal for a chemical memoir in 1908 and com-pleting a doctoral dissertation on Newton's me-chanics in 1911.

He became increasingly interested in the history and philosophy of science and dedicated the rest

of his life to establishing the history of science as an independent discipline. In 1912 he founded *Isis*, a scholarly journal for this new field, and con-tinued to be its editor through 1952. After only five issues, publication was interrupted by World War I, but it was resumed in 1919. Meanwhile Sarton had fled Belgium and moved to the United States in 1915. As a research associate of the Car-negie Institution, from 1918 on, he used Harvard's Widener Library and offered a course in the history of science at Harvard, where he became a profes-sor in 1940. The History of Science Society was formed in the United States in 1924; it provided a professional organization for the new discipline and guaranteed the publication of *Isis*. *Isis* has pub-lished a continuous series of valuable critical bibli-ographies; the first 79 of these were prepared by Sarton himself. *Osiris* was founded in 1936 to ac-commodate studies that were too long for *Isis* but too short for separate publication.

When Sarton was invited, in 1916, to deliver the Lowell lectures, he chose Leonardo da Vinci as his subject. On this occasion he was struck with how little was known about ancient and medieval sci-ence leading up to Leonardo, and he decided that a comprehensive history needed to be written. In 1919 he recovered his notes and a good part of his library, which he had had to abandon in Belgium, and two years later he began to write his *Introduc-tion to the History of Science* (1927–1948). Ulti-mately, Sarton did not take this history beyond the fourteenth century; he realized how much time would be required to handle the fifteenth century on the same scale as the earlier centuries and pre-ferred to spend his remaining years on more lim-ited subjects; for example, he wrote *A Guide to the History of Science* (1952a), *The Appreciation of Ancient and Medieval Science During the Renais-sance* (1955), and *Six Wings: Men of Science in the Renaissance* (1957).

His aim in writing the history of science was not only to establish and define that field of study but also to show its importance for the philosophy of science and for any positive philosophy. Sarton called science "systematized positive knowledge," picturing it as a living, continually developing or-ganism. He maintained that science is the only cumulative and progressive human activity and that the study of the history of mankind must focus on the history of science; any tolerably complete history of civilization must devote considerable space to explaining scientific progress. He believed that the historian of science has a mission to de-fend the traditions of science, not because they differ essentially from those of other fields, but



because "they are all that makes life worth living, they are the nobility and the goodness of life" (1952a, p. 15).

For Sarton, science has grown "out of the minds and hearts of living men" (1954, p. 67), and its history is therefore often best written in terms of biography. He did not limit his account to the achievements of outstanding scientists, because he found that a number of smaller discoveries usually contributed to the gradual preparation of major ones. He warned that the history of science must not be confused with the history of scientific method, nor be approached from the economic point of view, which overemphasizes applications.

Influenced by Comte and Tannery, Sarton repeatedly emphasized that the history of science is the integration of the histories of the separate sciences, not their summation; it is itself a specialty built on a thorough understanding of the methods of science and of history; it covers ancient, medieval, and modern science; and it requires more than the leisure hours of capable scientists or of scholarly historians. The historian of science must be trained in science; indeed, he should be acquainted with all advancing knowledge, not just scientific and historical material. Moreover, Western science cannot be properly understood without an appreciation of the extent of the influence of Islam on Western civilization, and Sarton learned Arabic in order to comprehend this influence.

Sarton's history of science, an all-embracing picture of the impact of scientific progress on mankind and of man's environment on science, tells "not only of the spreading light but also of the contracting darkness" (1952a, p. 11). Sarton saw the history of science as a means of understanding science and making it live—creating a philosophy of science that would have continuing validity. He combated the tendency of scientific studies to supplant humanistic ones. He strove to "humanize" science, to reveal its integration with the other elements of our culture, and to instill in scientists some reverence of the past and some awareness of the relevance of the history of science for science as an ongoing activity. The history of science seemed to him the only possible bridge across a growing abyss.

Warm and subtly humorous, Sarton made friends throughout the world. He was tireless in his own work and expected similar patience and critical attitudes from others. Sometimes his genuine humility seemed like self-righteousness (e.g., [1931] 1956, p. 109). His lecture courses were popular, but he had few graduate students.

A new generation of historians of science, while appreciative of Sarton's pioneering service, is reassessing his aims and methods. Present-day emphasis is on the structure and development of scientific ideas rather than on the ideas themselves. Sarton has been described as naive about history (Williams 1960, p. 160) and insufficiently selective in his methods, too lost in pedantic minutiae to appreciate their relevance for the overall development. He has been accused by some of stressing the content of science at the expense of its integration into a complex whole. On the other hand, Sarton's interpretation of the history of science is alleged to be incompatible with the compartmentalism of modern science and to stress the cultural framework and the normative implications of science at the expense of its content (Grmek 1965, pp. 140–143). Grmek questioned whether one man can handle the history of science if it is interpreted as broadly as it was by Sarton, but he admitted that collaborative studies by teams of specialists have gained technical insight at the expense of the kind of synthesis that Sarton provided. Clearly, Sarton's critics are in disagreement. However, the monumental scholarly edifice he erected can become a basis for a revision of the historiography of science.

C. DORIS HELLMAN

[See also SCIENCE, article on THE HISTORY OF SCIENCE.]

#### WORKS BY SARTON

- 1927–1948 *Introduction to the History of Science*. 3 vols. Baltimore: Williams & Wilkins.
- (1931) 1956 *The History of Science and the New Humanism*. 3d ed. New York: Braziller. → A paperback edition was published in 1962 by the Indiana University Press.
- 1952a *A Guide to the History of Science: A First Guide for the Study of the History of Science, With Introductory Essays on Science and Tradition*. New York: Ronald Press.
- 1952b *A History of Science. Ancient Science Through the Golden Age of Greece*. Cambridge, Mass.: Harvard Univ. Press.
- 1954 *The Old World and the New Humanism*. Pages 63–68 in *Man's Right to Knowledge; First Series: Tradition and Change*. New York: Muschel.
- 1955 *The Appreciation of Ancient and Medieval Science During the Renaissance (1450–1600)*. Philadelphia: Univ. of Pennsylvania Press.
- 1957 *Six Wings: Men of Science in the Renaissance*. Bloomington: Indiana Univ. Press. → Published posthumously.
- 1959 *A History of Science: Hellenistic Science and Culture in the Last Three Centuries B.C.* Cambridge, Mass.: Harvard Univ. Press.
- On the History of Science: Essays*. Selected and edited by Dorothy Stimson. Cambridge, Mass.: Harvard Univ. Press, 1962.

## SUPPLEMENTARY BIBLIOGRAPHY

- AGASSI, JOSEPH 1963 *Towards an Historiography of Science. History and Theory*, Beiheft 2. The Hague: Mouton.
- BUCHDAHL, GERT 1965 A Revolution in Historiography of Science. *History of Science* 4:55-69.
- The George Sarton Memorial Issue. 1957 *Isis* 48, part 3.
- GRMEK, MIRKO D. 1965 Prolégomènes à une histoire générale des sciences. *Annales: Économies, sociétés, civilisations* 20:138-146.
- HELLMAN, C. DORIS 1958 George Sarton: Historian of Science and New Humanist. *Science* 128:641-644.
- KUHN, THOMAS S. 1962 *The Structure of Scientific Revolutions*. Univ. of Chicago Press. → A paperback edition was published in 1964.
- STRELSKY, KATHARINE 1957 Bibliography of the Publications of George Sarton. *Isis* 48:336-350.
- WILLIAMS, L. PEARCE 1960 [A Book Review of] *A History of Science*, Vol. 2, by George Sarton. *British Journal for the Philosophy of Science* 11:159-161.

## SAUER, CARL O.

Carl O. Sauer, for more than thirty years chairman of the department of geography at the University of California, Berkeley, is one of the most influential figures in twentieth-century geography. The study of cultural geography, especially as it has developed in the United States, owes more to Sauer than to any other man. His concern for the quality of human life in its physical setting, for historical and ecological processes, and for the origins and dispersals of culture traits and complexes that have modified the cultural landscape is in the tradition of German geographical thought, of which Sauer has been a principal interpreter.

Born of German ancestry in Warrenton, Missouri, in 1889, he graduated from the local and now defunct Central Wesleyan College in 1908, having earlier studied for a time in a school in Württemberg, Germany, where he had family ties. His attachment to the rural and small-town ways of life and landscape of the Middle West of that period has played an important role in his teaching and writing, and he has frequently returned there for study and reflection.

Sauer next studied at the University of Chicago, where the first full-fledged graduate program in geography in the United States had recently been instituted. His PH.D. dissertation, *The Geography of the Ozark Highland of Missouri* (1920), was acclaimed as a model of regional cultural geography. In 1915 he was appointed to the staff of the new department of geology and geography at the University of Michigan, where in a period of seven years he advanced from the rank of instructor to that of full professor. His work in designing the

Michigan Land Economic Survey brought him national recognition. He left Ann Arbor in 1923, at the age of 33, to become chairman of the department of geography at Berkeley. Sauer retired from active teaching in 1957, except for occasional graduate seminars and guest lectureships, but his scholarly output has been undiminished.

Sauer also served the development of geography outside the university. He played an important role in the early development of land-use mapping in the United States and in the establishment of the U.S. Soil Conservation Service. For a time in the 1930s he was consultant to the President's Science Advisory Board. For many years he served as a member of the John Simon Guggenheim Memorial Foundation selection committee, in which capacity he had a unique opportunity to observe and influence a wide range of American scholarship. He has received honorary degrees from Heidelberg University, Syracuse University, and the University of California and medals from the American Geographical Society, the Berlin Geographical Society, and the Swedish Society for Anthropology and Geography.

Soon after Sauer came to California, he and many of his students turned their attention to northern Mexico, especially to the historical geography of the settlement of the dry northwestern part. Later he was to concern himself increasingly with tropical Mexico and Central America and, finally, with the Caribbean. Latin America was thus the strategic site for his observation of the interaction between the simpler, peasant cultures and the sustaining earth. Here, where the mark of the past is so deeply imprinted on the land, he found an especially congenial atmosphere for the study of the interplay between habitat and habitant. His careful reconstructions of past Latin American landscapes and economies, based on a combination of archival work and detailed field inspection, are models of historical geography. *The Early Spanish Main* (1966), an analysis of the Spanish conquest on the islands of the Caribbean and its margins, is their capstone. Several of his publications on Latin American subjects appeared in the University of California monograph series *Ibero-Americana*, which he founded in 1932 with A. L. Kroeber.

Sauer's outspoken rejection of environmental "influences" as the focus of geographical inquiry, enunciated in his *Morphology of Landscape* (1925), was a critical turning point in the development of American academic geography. Instead of the deterministic view that had largely dominated Ameri-



can geography since the turn of the century, he advocated geographical inquiry into such themes as the man-modified landscape and the origin, spread, limits, and alteration of the elements of that landscape. Holding that all geography is essentially historical, he systematically used any kind of data pertaining to human settlement and activity, and he often found archeological evidence fully as pertinent as written documents. For him the kitchen midden, the colonial archive, the relict plant community or soil profile, and the wisdom of the countryman at work on his hillside garden patch provided information equally relevant to the problems of cultural geography.

A recurring theme in Sauer's work is the tremendous importance and antiquity of the domestication of plants and animals. In *Agricultural Origins and Dispersals* (1952a), he summarized his efforts to locate the principal hearths of domestication and the subsequent routes and extent of dispersal. Using archeological and ethnographic evidence, Sauer argued that the origins of cultivation must have been in the woodlands rather than in riverine valleys of the world. He made a basic distinction between the agricultural complexes of monsoon Asia and northwestern South America, which are largely dependent on vegetative reproduction by slips and cuttings, and the seed cultures of southwest Asia and of Mexico and Central America. Sauer hypothesized that the earliest form of plant selection involved asexual plant propagation and that it was probably first practiced by the sedentary fishing peoples living around the shallow seas of southeast Asia. The early domestic animals of these peoples were household animals such as the dog, pig, and chicken. Sauer accepted the view that plant selection by seed reproduction came later, and he was influenced by N. I. Vavilov in pointing to China, India, and Ethiopia as Old World centers of seed domestication. He associated herd animals with cereal cultivation and accepted Eduard Hahn's theory of a mixed economy in which milking of animals was a culture trait integrally involved in the process and purpose of their domestication, probably for noneconomic ends. His research along these lines has especially attracted the attention of anthropologists and archeologists. Many of his speculations regarding the antiquity of man in the New World, Pleistocene ecology, and the relation of changing sea levels and shore configurations to human occupation sites continue to stimulate new research.

Sauer has insisted on the close links of cultural geography with both biology and the earth sciences. He sees man as the ecologic dominant in nature,

intervening more and more decisively to change the balance and nature of organic life. He has, for example, insisted that the problems of increasing world-wide desiccation are caused by agricultural and grazing practices and not by "climate deterioration." He has repeatedly written about the historical processes of biotic destruction and soil depletion as reflections of changing technology and changing attitudes about the earth. Another theme in his writings has been the role of fire, especially of fires set by man, in the modification of the vegetative covering of the earth and the establishment of grasslands. He was the principal organizer of an international symposium held at Princeton in 1955 that focused on "Man's Role in Changing the Face of the Earth" (1956a).

Sauer has viewed with misgivings the growing emphasis of some of his colleagues on techniques and tools of analysis that produce quantitative results: he has insisted that the qualities of both human society and the man-made landscape are often other than those that permit quantitative processing (1952b). Leighly has observed that to Sauer the investigation of human beings and their behavior by mechanical processes of organization and computation is as repugnant as the environmental determinism he rejected years earlier (see the introduction by Leighly, to Sauer's *Land and Life*, p. 6). Human geography, Sauer has said, exists because of the diversification that exists in human ways, and no approach to its questions is more rewarding than that of comparative culture history, soundly based on ecological and geographical principles and concepts.

JAMES J. PARSONS

[See also DOMESTICATION; ECOLOGY; GEOGRAPHY, especially the articles on CULTURAL GEOGRAPHY and SOCIAL GEOGRAPHY.]

#### WORKS BY SAUER

- 1920 *The Geography of the Ozark Highland of Missouri*. Geographic Society of Chicago, Bulletin No. 7. Univ. of Chicago Press.
- 1925 *The Morphology of Landscape*. University of California Publications in Geography, Vol. 2, No. 2. Berkeley: Univ. of California Press.
- 1932 *The Road to Cibola*. Berkeley and Los Angeles: Univ. of California Press.
- 1935 *Aboriginal Population of Northwestern Mexico*. Berkeley: Univ. of California Press.
- 1938 Theme of Plant and Animal Destruction in Economic History. *Journal of Farm Economics* 20:765-775.
- 1941 Foreword to Historical Geography. Association of American Geographers, *Annals* 31:1-24.
- 1948a *Colima of New Spain in the Sixteenth Century*. Berkeley: Univ. of California Press.

- 1948b Environment and Culture During the Last Deglaciation. American Philosophical Society, *Proceedings* 92:65-77.
- 1950 Grassland Climax, Fire, and Man. *Journal of Range Management* 3:16-21.
- 1952a *Agricultural Origins and Dispersals*. New York: American Geographical Society.
- 1952b Folkways of Social Science. Pages 100-109 in Minnesota, University of, Social Science Research Center, *The Social Sciences at Mid-century*. Minneapolis: Univ. of Minnesota Press.
- 1956a The Agency of Man on the Earth. Pages 49-69 in International Symposium on Man's Role in Changing the Face of the Earth, Princeton, N.J., 1955, *Man's Role in Changing the Face of the Earth*. Univ. of Chicago Press.
- 1956b The Education of a Geographer. Association of American Geographers, *Annals* 46:287-299.
- 1957 The End of the Ice Age and Its Witnesses. *Geographical Review* 47:29-43.
- 1958 Man in the Ecology of Tropical America. Pages 104-110 in Pacific Science Congress, Ninth, Bangkok, 1957, *Proceedings*. Volume 20: Special Symposium on Climate, Vegetation, and Rational Land Utilization in the Humid Tropics. Bangkok: The Secretariat.
- 1962a Homestead and Community on the Middle Border. *Landscape: Magazine of Human Geography* 12, no. 1:3-7.
- 1962b Seashore—Primitive Home of Man? American Philosophical Society, *Proceedings*. 106:41-47.
- 1966 *The Early Spanish Main*. Berkeley and Los Angeles: Univ. of California Press.
- Land and Life: A Selection from the Writings of Carl Ortwin Sauer*. Edited by John Leighly. Berkeley: Univ. of California Press, 1963.

#### SUPPLEMENTARY BIBLIOGRAPHY

- HAHN, EDUARD 1896 *Die Haustiere und ihre Beziehungen zur Wirtschaft des Menschen: Eine geographische Studie*. Leipzig: Duncker & Humblot.
- HAHN, EDUARD 1909 *Die Entstehung der Pflugkultur*. Heidelberg (Germany): Winter.
- HAHN, EDUARD (1914) 1919 *Von der Hacke zum Pflug*. 2d ed. Leipzig: Quelle & Meyer.

#### SAUSSURE, FERDINAND DE

Ferdinand de Saussure (1857-1913), Swiss linguist, was the chief forerunner of structural linguistics. He was born in Geneva, where he received his secondary and university education. He was awarded his doctorate at Leipzig in 1880, taught at the École Pratique des Hautes Études in Paris from 1881 to 1891, and taught subsequently at Geneva for the remainder of his career.

De Saussure's fame rests almost exclusively on two works—his earliest, the *Mémoire sur le système primitif des voyelles dans les langues indo-européennes* (1879), and the posthumously published *Cours de linguistique générale* (1916). The *Mémoire* is a brilliant tour de force, whose basic method foreshadows the fundamental structuralist notion of language as an organized system, the

central doctrine of de Saussure's mature thought. [See LINGUISTICS, article on THE FIELD.]

In the *Mémoire* it is boldly hypothesized that certain apparently divergent aspects of the Indo-European system of vowel alternations (the so-called ablaut system, which still survives in English in such vowel changes as the present *sing*, the preterite *sang*, the past participle *sung*) can be explained in accordance with the dominant regularities, if one assumes that certain sounds had formerly existed and produced certain effects before being lost. While this theory of lost sounds, subsequently called laryngeals, was admired in de Saussure's lifetime, the absence of any direct evidence for these sounds prevented its acceptance by scholars. After de Saussure's death the main points of his doctrine regarding the Indo-European vowels received confirmation when a newly discovered Indo-European language, cuneiform Hittite, was shown to have an actual laryngeal consonant (h) in almost all the instances that one would expect to find it on the basis of de Saussure's theory. In one variant or another, the laryngeal theory is now accepted by all Indo-European scholars.

A far wider influence was exercised by the *Cours*. This work is based on students' notebooks of the three courses in general linguistics given by de Saussure at the University of Geneva in 1907, 1908-1909, and 1910-1911. Two of de Saussure's leading disciples, Charles Bally and Albert Sechehaye, utilized these as their chief source in putting together the *Cours*. Given these circumstances of the book's composition, it is not surprising that an extensive exegetical literature has arisen concerning the true nature of de Saussure's doctrine. Godel (1957) published a considerable number of résumés and direct quotations from the notebooks and other unpublished sources, e.g., de Saussure's class notes and letters. It is quite likely that in some instances "fruitful misunderstandings" of de Saussure's ideas have exerted an influence on the development of European linguistic theory.

**Language as signs.** Language is a system of signs and as such forms the central part of a discipline—semiology—which takes the general study of signs as its subject matter. De Saussure thus anticipated the concept of general semiotics, found in the work of Charles Morris and others [see SEMANTICS AND SEMIOTICS]. The sign has two indissoluble aspects, likened to the two sides of the same paper—the signifier (*signifiant*) and the signified (*signifié*). De Saussure felt the need for the double terminology because of the ambiguity of the word "sign" in common usage, which refers



both to the relation (between a sign and what it designates) and to the physical sign vehicle. The signifier and the signified are, in thoroughgoing mentalistic fashion, the acoustic image and the concept respectively, not the physical sound and the thing meant. The relation between signifier and signified is arbitrary in the sense that any signifier can, in principle, be connected with any signified. Both of these elements are, abstracted from their functioning in signs, unstructured. Language is thus pure form.

Although arbitrary in the sense described above, the sign relation in any particular case is a historically imposed one about which the individual has no choice. It is a social fact that in a given community a particular sequence of sounds is connected with a particular concept. Simply because it is arbitrary, there is no reason to change it once established. One can argue the relative merits of monogamy or polygamy, but there is no rational basis for choosing between *boeuf* and *Ochse* as designations for "ox." Because language is arbitrary and because there is no social awareness of its true nature, it is never subject to radical change.

From the formal nature of the components of the linguistic sign further important conclusions follow. Thus, in the case of the signifier, it is not the particular physical nature of the sound that counts but merely the fact that one sound is distinct from another. A favorite Saussurean simile likens language to a chess game. If the knight is replaced by a different kind of physical object it makes no difference to the game as long as the object is distinct from the other pieces. But if the rules for moving the knight are modified, then the system of chess has been modified in its very nature. This distinction is one aspect of a central Saussurean dichotomy, that between language (*langue*) and speech (*parole*). Phonetics in its physical aspect belongs to the latter. Likewise in grammar, we have a network of two types of relations, the syntagmatic, the relation between elements which actually occur together in the chain of speech—e.g., a verb and its object—and the paradigmatic, the relation of a present element to one which is not there but has the potentiality of appearance in place of the one chosen. Thus, in the sentence "Black is a color," "white" is a paradigmatic associate of "black," while "is a" constitutes a syntagmatic relation.

The grammar of a language (*langue*) is the set of these two types of relations: it is a social product, never perfect in the individual but only in the mass of speakers (*masse parlante*). The way individuals actually construct sentences by using

*langue* constitutes *parole*. Within *langue* it is the position within the system that counts, not the physical realization. Thus a grammatical category may even be signaled by nothing (*zéro*), provided it has a particular place in a network of relations, e.g., the absence of an ending in the genitive plural of feminine nouns in Czech such as *žen* ("of women") is significant when it is opposed to the nominative plural, *ženy*. To be is to be related.

**Synchronic and diachronic study of language.** Another dichotomy of fundamental importance in de Saussure's thought is that between the synchronic study of language as a state and the diachronic study of language as change through time [see LINGUISTICS, article on HISTORICAL LINGUISTICS]. These now current terms are terminological inventions of de Saussure. The insistence on the separation of synchronic and diachronic and the emphasis on synchronic states, not only as worthy autonomous subjects for investigation but even as having a certain priority, were perhaps de Saussure's most revolutionary ideas. Linguistics in its recognizably modern form arose in the early nineteenth century as a historical discipline, and de Saussure was the first to state clearly the viewpoint of modern structuralism—that language studied in a single time plane constitutes a systematic set of relations that can be studied without reference to the historic process by which these have come to be what they are. Thus, to revert to de Saussure's figure of the chess game, the man who has seen the entire game up to a particular point has no advantage over the newly arrived bystander in understanding the existing position.

In spite of differences in detail, structuralism in linguistics is fundamentally allied with functionalism in social science generally and is part of the same intellectual movement. In the case of de Saussure, his notion of *langue*, a social phenomenon imposed on the individual, brings him closest to the social functionalism of Durkheim. Although Durkheim cited language as an ideal example of a social institution, there is apparently no evidence of any direct influence of either man on the other.

The posthumous influence of de Saussure's ideas in general linguistics was enormous, particularly on the Continent. His work was most directly continued by the so-called Geneva school, which included among its leading figures Albert Sechehaye, Charles Bally, Robert Godel, and Henri Frei and which sponsored the journal *Cahiers Ferdinand de Saussure* (from 1941 to the present). The glossematic approach developed by the Danish linguist Louis Hjelmsler was founded on Saussurean con-

cepts. De Saussure exercised a more diffuse influence on the chief European structuralist school, that of Prague, by which he was looked upon as a leading pioneer of structuralism in linguistics.

JOSEPH H. GREENBERG

#### WORKS BY SAUSSURE

- 1879 *Mémoire sur le système primitif des voyelles dans les langues indo-européennes*. Leipzig: Teubner.  
(1916) 1959 *Course in General Linguistics*. New York: Philosophical Library. → First published posthumously as *Cours de linguistique générale*.

#### SUPPLEMENTARY BIBLIOGRAPHY

- BENVENISTE, É. 1962 "Structure" en linguistique. Pages 31-39 in *Colloque sur le mot structure*, Paris, 1959, *Sens et usages du terme structure dans les sciences humaines et sociales*. Janua linguarum, Series minor, no. 16. The Hague: Mouton.  
*Cahiers Ferdinand de Saussure*. → Published in Geneva by Droz since 1941.  
FREI, HENRI 1947 *La linguistique saussurienne à Genève depuis 1939*. *Word* 3:107-109.  
GODEL, ROBERT 1957 *Les sources manuscrites du Cours de linguistique générale de F. de Saussure*. Geneva: Droz.  
WELLS, RULON S. 1947 *De Saussure's System of Linguistics*. *Word* 3:1-31.

### SAVIGNY, FRIEDRICH CARL VON

Friedrich Carl von Savigny (1779-1861) was the founder of the older, romantic branch of the German historical school of law. He was born in Frankfurt am Main of an ancient and wealthy noble family from Lorraine, and his selection of an academic career was unusual for a member of his class. He began that career at the University of Marburg in 1800, teaching Roman law, and became professor of Roman law at Landshut in 1808. Two years later, at the suggestion of Wilhelm von Humboldt, he was appointed professor at the newly founded University of Berlin, where for three decades he had a most successful and influential career in legal teaching and research. From 1842 until the German revolution of 1848 he was Prussian minister for the revision of legislation.

In the academic year 1802/1803, Savigny gave a lecture on legal methodology at Marburg (see *Juristische Methodenlehre*), which shows him still poised between rationalistic and historical jurisprudence. (His student Jacob Grimm recorded this lecture.) However, the lecture contains some of the basic principles of his later doctrine, for example, that jurisprudence must be empirical and that it should be a synthesis of historical and systematic treatments of the law. He first applied these early

theories in *Das Recht des Besitzes* (see 1803), a treatise that immediately established his reputation as one of the leading professors of civil law. What contemporary legal scholars saw as the particular significance of the treatise was the way Savigny took from the sources of ancient Roman law the leading principles of the law of possession, demonstrating the superiority of the Roman principles over the modifications that this law had undergone in Continental jurisprudence since the Middle Ages.

Savigny's famous pamphlet *Vom Beruf unserer Zeit für Gesetzgebung und Rechtswissenschaft* (1814) was published as an argument against the popular demand for the codification of German civil law on the model of the French and the Austrian codes. The pamphlet established the doctrine and program of the early historical school of law. In this pamphlet, as well as in his introductory essay (1815) to the first volume of the *Zeitschrift für geschichtliche Rechtswissenschaft*, founded as the organ of the school in 1815, Savigny outlined a new theory of the creation of law, culminating in a strictly genetic conception and analysis of law. Taking off from the ideas of Montesquieu, Burke, and Justus Möser, he asserted that positive law is not a product of reason. From its true basis, the common beliefs of a particular nation, it develops in a continuous, "organic" process as do a nation's language and customs; it is the result of silent, anonymous, and irrational forces, of an implicit consensus, not of an intentional creative effort. Law is, therefore, of the same nature as custom; it is determined by the whole past of the nation and by its peculiar character (what was later called the *Volksgeist*), and it cannot be changed arbitrarily by legislation. Instead, Savigny advocated the reform of jurisprudence by means of historical research, holding that a strictly genetic study is the only way to conceive and develop contemporary positive law.

While some of Savigny's contemporaries (Thibaut, P. J. A. von Feuerbach, Gönner, and especially Hegel and his school) rejected his doctrine, the historical school of jurisprudence dominated German universities for nearly half a century. Its program of historical research, though neither wholly nor rigorously fulfilled, made the history of law a scientific discipline, by basing it on the idea of evolution. Savigny's own chief contribution in this field was his *Geschichte des römischen Rechts im Mittelalter* (1815-1831), which deals with the history of the sources, the literature, and the teaching of Roman law from the fifth through the fifteenth centuries. It initiated the modern study of medieval jurisprudence and has served ever



since, corrected in many respects but unequalled, as the standard work on the history of medieval jurisprudence in Europe. Savigny also wrote a number of historical monographs, mainly on ancient Roman law (collected in his *Vermischte Schriften*).

In the field of legal dogmatics, Savigny's doctrine re-created, by reviving the study of ancient Roman law, a common field of legal study in the individual German states—the "modern Roman law," or Pandects. His most important dogmatic works, *System des heutigen römischen Rechts* (1840–1849) and *Das Obligationenrecht als Theil des heutigen römischen Rechts* (1851–1853), show that the theories that he had put forward in 1814 and 1815 about the nature of law had not produced a radical change even in his own legal dogmatics. His early theories had been provoked by his reaction, influenced by romanticism, against the natural law theories of the French Revolution and of liberalism generally. Later, however, he not only retained the fundamental doctrines of the rationalist jurisprudence of the eighteenth century in his dogmatic teaching but actually made these doctrines the backbone of his legal system. Both of Savigny's works are based on fundamental notions of German natural law theory as well as on Kant's legal philosophy. Within the framework of these categories, Savigny transcended Roman law in many respects, adapting it to modern requirements or developing new solutions, many of which have become influential in German and other legal science; for example, his doctrines of contract and corporation, and particularly his system of private international law—which, like the systems of Story and Mancini, counts among the classic theories in this branch of law.

WALTER WILHELM

[For the historical context of Savigny's work, see CONFLICT OF LAWS; JURISPRUDENCE; LEGAL SYSTEMS, article on CODE LAW SYSTEMS; and the biographies of BURKE; MONTESQUIEU.]

#### WORKS BY SAVIGNY

- (1803) 1848 *Von Savigny's Treatise on Possession: Or, the Jus Possessionis of the Civil Law*. 6th ed. London: Sweet. → First published as *Das Recht des Besitzes*. A 7th German edition, enlarged and revised, was published in 1865 by Heyer.
- (1814) 1840 *Vom Beruf unserer Zeit für Gesetzgebung und Rechtswissenschaft*. Heidelberg: Mohr. → Translated from the 2d edition by A. Hayward and published by Littlewood in 1831 as *Of the Vocation of Our Age for Legislation and Jurisprudence*.
- (1815) 1850 *Ueber den Zweck der Zeitschrift für geschichtliche Rechtswissenschaft*. Volume 1, pages 105–126 in Friedrich Carl von Savigny, *Vermischte*

- Schriften*. Berlin: Veit. → First published in Volume 1 of *Zeitschrift für geschichtliche Rechtswissenschaft*. (1815–1831) 1834–1851 *Geschichte des römischen Rechts im Mittelalter*. 7 vols. Heidelberg: Mohr. → Volume 1 was translated by E. Cathcart and published by Black in 1829 as *The History of the Roman Law in the Middle Ages*.
- 1840–1849 *System des heutigen römischen Rechts*. 8 vols. Berlin: Veit. → Volume 1 was translated in 1867 by William Holloway as *System of the Modern Roman Law*, published in Madras, India, by Higginbotham. An index volume, edited by O. L. Heuser, was published in 1851.
- 1851–1853 *Das Obligationenrecht als Theil des heutigen römischen Rechts*. 2 vols. Berlin: Veit.
- Juristische Methodenlehre, nach der Ausarbeitung des Jakob Grimm*. Edited by Gerhard Wesenberg. Stuttgart: Koehler, 1951. → Published posthumously.
- Vermischte Schriften*. 5 vols. Berlin: Veit, 1850. → Contains works first published between 1800 and 1844.

#### SUPPLEMENTARY BIBLIOGRAPHY

- GAGNÉ, STEN 1960 *Studien zur Ideengeschichte der Gesetzgebung*. Uppsala, Universität, Acta universitatis upsallensis, Studia iuridica upsallensis, No. 1. Stockholm: Almqvist & Wiksell. → See especially pages 15–51.
- GUTZWILLER, MAX 1923 *Der Einfluss Savignys auf die Entwicklung des Internationalprivatrechts*. Collectanea friburgensia, Vol. 28. Fribourg (Switzerland): Universitäts Buchhandlung.
- KANTOROWICZ, HERMANN 1912 *Was ist uns Savigny?* Berlin: Heymann.
- KANTOROWICZ, HERMANN 1934 Savigny, Friedrich Carl von. Volume 13, pages 546–548 in *Encyclopaedia of the Social Sciences*. New York: Macmillan.
- KANTOROWICZ, HERMANN 1937 Savigny and the Historical School of Law. *Law Quarterly Review* 53:326–343.
- KOSCHAKER, PAUL (1947) 1966 *Europa und das römische Recht*. 4th ed. Munich: Beck.
- KUNKEL, WOLFGANG 1962 Savignys Bedeutung für die deutsche Rechtswissenschaft und das deutsche Recht. *Juristenzeitung* 17:457–463.
- STINTZING, RODERICH VON; and LANDSBERG, ERNST (1910) 1957 *Geschichte der deutschen Rechtswissenschaft*. Volume 3, part 2: Das 19. Jahrhundert, der Historismus in der Rechtswissenschaft. Aalen (Germany): Scientia Antiquariat. → See especially pages 186–253 of the text section, and pages 94–110 of the notes.
- STOLL, ADOLF 1927–1939 *Friedrich Carl v. Savigny: Ein Bild seines Lebens mit einer Sammlung seiner Briefe*. 3 vols. Berlin: Heymann.
- WIEACKER, FRANZ 1956 Friedrich Carl von Savigny. *Zeitschrift der Savigny-Stiftung für Rechtsgeschichte Romanistische Abteilung* 72:1–38. → Reprinted on pages 107–143 in Franz Wieacker, *Gründer und Bewahrer: Rechtslehrer der neueren deutschen Privatrechtsgeschichte*, published by Vandenhoeck and Ruprecht, 1959.
- WOLF, ERIK (1939) 1963 *Grosse Rechtsdenker der deutschen Geistesgeschichte*. 4th ed. Tübingen (Germany): Mohr. → See especially pages 467–542.

#### SAVING

See CONSUMPTION FUNCTION; INTEREST; INVESTMENT.

## SAY, JEAN BAPTISTE

Jean Baptiste Say (1767–1832) was born in Lyons of a cultivated Protestant merchant family. After a sound education and a two-year stay in England, he joined an insurance company. Say was an enthusiastic supporter of the ideas of the French Revolution and took part as a volunteer in the military campaign of 1792. Bookish and inclined to philosophy, in 1793 he joined the editorial board of a new magazine entitled *La décade philosophique littéraire et politique, par une société de républicains*, which took as its motto "Enlightenment and morality are as much needed for preserving the Republic as courage was needed for winning it." He soon became the editor-in-chief of the journal, which announced that its aim was "to keep the public informed on everything new concerning literature, science, agriculture, and the arts."

In 1799 Say was appointed a member of the Tribunal, under the Consulate. He took part in a competition arranged by the Institut de France on the subject "By what means and what institutions can morality be established in a people?" His contribution, "Olbie: Ou, essai sur les moyens de réformer les mœurs d'une nation," gave promise of his gifts as a moralist and economist. Following up his thesis that "the first book on morality for the Olbians should be a good Treatise on Political Economy" (*Oeuvres diverses*, p. 593), he set about composing such a work, inspired to a great extent by the ideas expounded by Adam Smith in his celebrated *Inquiry Into the Nature and Causes of the Wealth of Nations*, with which Say had become acquainted as early as 1789. Say's book appeared in 1803 and soon had considerable success.

In the meantime, Say's disapproval of the Napoleonic regime led to his dismissal from the Tribunal. He moved to a small town in northern France, Auchy-les-Hesdin, to set up a cotton-spinning plant. He stayed there seven years before he sold out in 1813 and returned to Paris. The fall of the empire and the return of the Bourbons permitted him to resume intellectual activity aimed at the general public. In 1815 he began to teach at the Athénée what was probably the first public course of political economy ever given in France, and in the same year he published his *Catéchisme d'économie politique*. Two years later the government established a chair of industrial economy for him (the term political economy still frightened people) at the Conservatoire des Arts et Métiers. Finally, in 1830, the aged master did become professor of political economy at the Collège de France, in the first chair of its kind to be established. Meanwhile, he had

published his *Cours complet d'économie politique pratique* (1828–1829) in six volumes, a work, as the long subtitle indicates, "designed to teach statesmen, landowners and capitalists, men of learning, agriculturalists, manufacturers, businessmen, and in general all citizens, about political economy."

Although it has been alleged that Say was a mere expositor and popularizer of Adam Smith's ideas, he did more than merely transmit these ideas. He did, of course, present the ideas that he borrowed from Smith with typical French order and clarity, and he justly complained of obscurity and confusion in Smith's own statements in *The Wealth of Nations*. Say wanted to produce what he described as "a work in which sound observations were reduced to general principles that could be accepted by any judicious man; in which those observations and principles were so completed and coordinated as to reinforce one another, so that they could fruitfully be studied at any time and in any place" (1803, p. xlviii in 1860 English edition). And this was not merely an interpretative work: at important points Say showed originality, and he undoubtedly contributed to advancing the science of economics.

**Philosophical position.** In his basic philosophy, Say was strongly influenced by the ideological ambience of his time. Like many of his contemporaries, he was at once deist, liberal, and rationalist. In the realm of scientific and doctrinal convictions, however, he showed little toleration but instead a dogmatism that may have been related to his authoritarian nature. This is one of several ways in which he differed sharply from Adam Smith: he was never touched by doubt. As a disciple of the Encyclopedists, he was an intransigent scientist, even a passionate one, and convinced of the predominant influence of economic phenomena in the life of the whole society. On this point, as in the case of the preponderant role of industrial activity in production as a whole, it may be possible to establish a connection between Say's ideas and certain theories of Karl Marx, and in any event, with those of Saint-Simon and the Saint-Simonians.

His belief in economics as a science led Say to try to reconcile his naturalism and his rationalism. He was able to do so because he conceived of the laws of nature as necessarily rational (or providential); in this respect, he was adopting a point of view the physiocrats had developed a half century earlier. But it is noteworthy that Say went on to establish similarities between the science of economics and *physiology*, transposing the latter from the individual to the social level and preferring a



biological model of the economy to a purely mechanical one. He was later to be followed on the same track by Alfred Marshall, the great English economist. It should be pointed out, however, that Say's method was never made rigorously explicit. While asserting the sovereignty of facts in scientific matters, he nonetheless interpreted them according to the results that he intended to reach. Thus, he appeared as an "ideologue" to some of his contemporaries.

**Relationship to Adam Smith.** Say deeply admired the work of Adam Smith. Of *The Wealth of Nations* he said, "Whenever the *Inquiry Into the Wealth of Nations* is perused with the attention it so well merits, it will be perceived that until the epoch of its publication the science of political economy did not exist" (1803, p. xxxviii in 1860 English edition). But Say was not content merely to repeat or rethink the ideas of his master. He filtered them, clarified them, and at several points went beyond them. He made a more systematic critique of physiocratic ideas, based on an expansion of the concept of productivity and the inclusion of immaterial products among the positive results of economic activity. Say gave greater depth to value theory by introducing the notions of need and utility. As a true precursor of the marginalist school, he noted that the needs of men may be classified according to their urgency in scales of preference, that they give rise to utility, which in turn determines demand. Demand then interacts with supply, which in turn is determined by the cost of production. For Say, however, in contrast to Ricardo, for example, utility is the stronger determining factor in prices; the cost of production enters only as a kind of restrictive condition. There too Say appears to have anticipated Alfred Marshall's well-known theory. It is of interest to see that Say realized perfectly that the explanation of the phenomena of value and price is not to be sought in terms of unilateral causal relationships but in terms of the concept of interdependence.

Even more than Adam Smith, Say strongly emphasized that the role of science is primarily one of observation and explanation. Where Smith, following the physiocrats, came forward as a physician for society, Say aimed exclusively at being a naturalist. His *Traité* professes to be a "simple exposition" of a body of exact knowledge, applicable under all circumstances.

Although agriculture still had a predominant position for Smith, it no longer did for Say. To be sure, the half-century separating their work witnessed the entire industrial revolution in Britain, and in France its beginning. Say was in a position

to observe its first manifestations, in the seven years that he devoted to business. Thus he made industry and even commerce as important as agriculture in the total productive activity of society.

**Major contributions.** Say appreciated the importance of the *entrepreneur* in economic life, a role Smith had neglected to define and which has remained, since Say's time, a permanent feature of economic analysis. This essential actor has functions that are quite distinct from those of the capitalist strictly so called (British authors confused these two conceptions). The entrepreneur, whether he be an industrialist, businessman, or farmer, has a truly strategic position in economic life, in that he is par excellence the intermediary between landowners, capitalists, workers, and consumers. He is also pictured as performing a truly creative function as the organizer responsible for production, striving unceasingly to improve it, especially in the industrial sector. Over a century later, Schumpeter was to take up and perfect this dynamic conception of the role of the true entrepreneur, the creator of new combinations (innovations).

Say's development of the theory of the distribution of the social product, deduced from the previous analysis, was a marked advance over current theories. Say coordinated this theory precisely and harmoniously with the theories of exchange and production. The entrepreneur, a buyer of productive services at known prices (the services of men, capital, and lands) and a seller of products at uncertain prices, is thus seen as assuming a set of fundamental functions, as presiding in a way over activity that is both production and distribution. His specific share is the profit, a share that is described as essential, residual, and precarious. This brings us to the very threshold of Walras's grand conception of economic equilibrium.

Another of Say's highly original contributions to economic analysis is his famous *law of markets*, which undoubtedly did more to make its author celebrated than all the rest of his work. This law (cf. Schumpeter in *History of Economic Analysis* 1954, pp. 615-626) may be interpreted in many different ways. We know its very simple, almost lapidary content: "Products are exchanged for products." This may be taken as meaning that the role of money in the economy is, after all, secondary (the conception of the "monetary veil"). But one may also infer that everyone taking part in an exchange is interested in the prosperity of his partner, whether it be an individual or a nation, a thesis that goes counter to the mercantilist conception that "one gains only what the other loses." Finally, the law may imply a theory of crises, based

on rejection of the concepts of general overproduction and underconsumption and the assertion of the essentially transitory and limited nature of these imbalances. The remedy is therefore not the limiting of production but its expansion. "Produce, produce, that is the whole thing!" Say concluded, in his *Cours complet*, that although "there is an excess in certain products, there is a shortage in others."

**Importance and influence.** All in all, it would appear that Say's work, without bearing the marks of really great genius, does give its author an eminent place in the development of economic thought. Say was not entirely an innovator, even in fields where he was not following Adam Smith and did show real originality: on many points he links up to a tradition solidly rooted in France. As Schumpeter remarked judiciously (1954, p. 492): "Say's work is the most important of the links in the chain that leads from Cantillon and Turgot to Walras."

He opened up new paths, but later authors followed them with more success than he. This was the case with the members of the marginalist school—Carl Menger, Stanley Jevons, and especially Léon Walras—who were able to employ the notion of utility in a much more precise and scientifically valid manner than their common precursor, Say. (Walras tended to minimize his debt to Say, but it was nevertheless important.) Again, the fine effort at synthesis made by Alfred Marshall and the neoclassical school is not entirely untouched by the teachings of Say. His influence can be traced beyond the general equilibrium school to such essentially modern authors as Schumpeter.

It is true that in the eyes of many of our contemporaries Say is seen primarily as the author of the law of markets, one of the favorite butts of Keynesian and Neo-Keynesian criticism, and this "law," interpreted and misinterpreted as it has been, may remain his chief title to fame. But perhaps he will be remembered for his power to build on established intellectual traditions and to stimulate other thinkers: there lay his true merit, which only time will confirm.

GASTON LEDUC

[For the historical context of Say's work, see ECONOMIC THOUGHT, article on PHYSIOCRATIC THOUGHT; and the biographies of SAINT-SIMON; SMITH, ADAM. For discussion of the subsequent development of his ideas, see DEMAND AND SUPPLY; ENTREPRENEURSHIP; UTILITY; and the biographies of JEVONS; MARSHALL; MENDER; SCHUMPETER; WALRAS.]

#### WORKS BY SAY

(1803) 1964 *A Treatise on Political Economy: Or, the Production, Distribution and Consumption of Wealth.*

New York: Kelley. → First published as *Traité d'économie politique: Ou, simple exposition de la manière dont se forment, se distribuent, et se consomment les richesses*. Translation of extract in text provided by the editors.

(1815) 1816 *Catechism of Political Economy: Or, Familiar Conversations on the Manner in Which Wealth Is Produced, Distributed, and Consumed in Society.* London: Sherwood. → First published as *Catéchisme d'économie politique*.

(1828–1829) 1852 *Cours complet d'économie politique pratique*. 2 vols. 3d ed., enl. With notes by Horace Say. Paris: Guillaumin. → First published in 6 volumes. Translation of extracts in text provided by the editors.

*Oeuvres diverses*. Paris: Guillaumin, 1848. → Translation of extract in text provided by the editors.

#### SUPPLEMENTARY BIBLIOGRAPHY

GOETZ-GIRY, ROBERT 1966 *Croissance et progrès à l'origine des sociétés industrielles*. Paris: Montchrestien.

SCHUMPETER, JOSEPH A. (1954) 1960 *History of Economic Analysis*. Edited by E. B. Schumpeter. New York: Oxford Univ. Press.

TEILHAC, ERNEST 1927 *L'oeuvre économique de Jean-Baptiste Say*. Paris: Sirey.

### SCALE, ECONOMIES OF

See ECONOMIES OF SCALE.

### SCALES OF MEASUREMENT

See PSYCHOMETRICS; STATISTICS, DESCRIPTIVE; SCALING.

## SCALING

Measurement of a property involves assignment of numbers to objects as a way of representing that property. The process thus includes both a formal, logical system, e.g., the real number system, and an empirical system, the set of instances of the property.

**Scale types.** An important aspect of measurement is the correspondence between some of the characteristics of the number system involved and some of the relations between instances of the property. The particular set of characteristics of the formal number system that is used determines the *type of scale* that is obtained. Scale types can most easily be described by the type of transformation that can be applied to a given scale without changing the numerical interpretation of the basic empirical relations. Although there is no limit to the number of scale types that might be devised, only a relatively small number have been found useful. Some of the more important scale types are described here.

In ordinal scales the only relevant property of numbers is their *order*. Numbers are assigned to



objects so that the order of the numbers corresponds to the order of the objects with respect to the property of interest. Any transformation of a given numerical assignment that preserves order, i.e., any *monotonic* transformation, would serve as well. An ordinal scale is therefore said to be determined up to a monotonic transformation.

*Fixed-origin ordinal* scales are rarely discussed, but are of some empirical importance in the fields of social psychology and personality, where bipolar attributes are not uncommon. Of relevance in such scales is the fact that numbers have a unique origin. Numbers are assigned not only so that their order corresponds to the order of the objects but also so that the origin of the numerical series corresponds to an empirically meaningful origin or zero point of the property. Any monotonic transformation that does not change the origin leaves the numerical interpretation of the basic relations unchanged.

In *interval* scales, in addition to order, differences between the numbers correspond, in some generalized sense, to the distances or differences between the instances of the property. Interval scales are thus determined up to a linear transformation of the form

$$y = ax + b,$$

where  $a$  is any positive number and  $b$  is any number whatsoever.

*Log interval* scales are closely related to the interval scale. In the interval scale, numerical differences correspond to empirical differences; in the log interval scale, numerical ratios correspond to empirical ratios. The transformation characterizing the log interval scale is given by

$$y = ax^b,$$

where  $a$  and  $b$  are positive.

*Ratio* scales can be described either as those in which the order, the differences, and the ratios of the assigned numbers correspond to the order, differences, and ratios of instances of the property or as those in which order, differences, and origin all receive numerical interpretation. For either description, the ratio scale is determined up to a linear transformation of the form  $y = ax$ , where  $a$  is positive.

*Difference* scales and *ratio* scales are related in a way similar to the way that interval and log interval scales are related; in each case one scale is equivalent to the logarithm of the other. In the difference scale, the procedures used to assign numbers serve to fix the order, the differences between numbers, and the unit of measurement. The char-

acterizing transformation is  $y = x + a$ . [For a more extensive discussion of scale types, see *PSYCHOMETRICS; STATISTICS, DESCRIPTIVE*; see also Suppes & Zinnes 1963.]

The concept of scale type is important because the type of scale on which a property is measured restricts the kinds of information that can be obtained in an experiment. The scale type therefore limits the kinds of meaningful questions that can be asked regarding the property. Consider the following three questions:

*Do high-scoring subjects increase in ability more than low-scoring subjects (after some specified treatment)?* If the ability is measured on a scale that contains only ordinal information, and if both low and high scorers increase in ability with no overlap in distribution, then the question cannot be answered. The question deals with the relative magnitude of differences (or perhaps ratios), and in a purely ordinal scale the relative magnitude of numerical differences or ratios is meaningless.

*Is amount of aggression directly proportional to degree of frustration (in some specified situation)?* Let  $A$  denote amount of aggression and  $F$  the degree of frustration. The question then concerns the form of the equation that relates  $A$  to  $F$ , that is, one asks whether these two variables are functionally related by an equation of the form  $A = kF$ , where  $k$  is some positive constant. Clearly, an affirmative answer cannot be given unless both  $A$  and  $F$  are measured on ratio scales.

*Do college sophomores differ more widely in their attitude toward civil rights than in their attitude toward the Asian policy of the United States?* Questions such as this require a scale type ordinarily ignored in physical measurement. Consider the formally equivalent question "Does this set of containers differ more widely in their volume than in their weight?" The question, which does not seem particularly unreasonable in the attitude context, now appears absurd: a difference of  $x$  cubic inches is neither more nor less nor equal to a difference of  $y$  pounds. The two differences are simply not comparable. However, volume and weight are measured on ratio scales. If they, or the two attitude variables, were measured on difference scales, then a basis for the comparison would exist.

*Kinds of measurement.* So far, scale type has been treated as though it were an inherent part of a given numerical assignment. In one sense, this is correct; each scaling procedure includes a set of rules governing assignment of numbers to instances of the property. The rules limit to varying extents and in varying ways the freedom of the investigator to decide which numbers to assign to the instances

of the property. The characterization of scale type by admissible transformations is simply a compact way of stating the particular restrictions imposed.

But scaling procedures also differ markedly in the quality of information the numbers represent. At one extreme are measurement procedures that rely almost entirely on definition or assumption. An experimenter might simply *define* the attitude of a subject to be equal to the number of favorable responses he gives to a specified collection of items. Such *measurement by fiat* has been both useful and common throughout the behavioral sciences. The type of scale resulting from such a procedure is simply a matter of *definition* or *assumption*. Although these procedures determine a unique number to be assigned to a given subject, the experimenter may have reasons for wanting to attribute meaning only to the ordinal property of the numerical assignment. Or he may want to attribute meaning to both ordinal and interval properties, thus obtaining an interval scale. Suitable modifications of the definition could therefore result in any of the scale types already described. Obviously, such modifications do not change the empirical basis of the numerical assignment, which, in the present instance, is simply the number of favorable responses to the particular set of items.

The usefulness of such defined scales depends heavily on the experience and intuition of the investigator. The scales tend to be exceedingly special. Different investigators, when defining scales in this way for the same attribute, would ordinarily select different, or at best only partially overlapping, sets of items. Because favorable responses depend as much on the nature of the items as on the attitude of the subject, the value assigned to a given subject would clearly be expected to differ from one scale to another. While the scales might be ratio scales by definition, it is unlikely that they would differ from each other only in unit of measurement. The procedures themselves provide no basis for choosing among the many scales of a given attribute that might be defined in this way. Hence, confidence in any one cannot be expected to be great.

A somewhat different procedure for obtaining scale properties by definition relies on assumptions made about the distribution of subjects. The investigator might stipulate, for example, that the number of favorable responses to his set of items only serves to order the subjects on the attribute. But he then might go on to assume that his population of subjects is distributed in some specified way on the attribute (ordinarily, the assumption of a normal distribution is made, but other distributions could serve as well). Knowledge of order and

of the properties of the assumed distribution enables the investigator to increase the level of the scale from ordinal to interval or perhaps even to ratio. But again, the meaning of the interval or ratio properties of the scale is derived only from the definition. There is no limit to the number of base populations that can be defined. And scales simply defined through use of the distribution assumption on one population cannot be expected to match those based on other populations.

A more appealing basis for establishing interval or ratio properties is provided by fundamental measurement procedures. In these procedures, the meaning of the properties of the final scale is determined, at least to some extent, by the empirical relations between the observables. Fundamental measurement procedures involve verifiable scientific models or theories, place restrictions on the behavior of the observables, and are subject to empirical test. One possible outcome of an application of a fundamental procedure, of course, is the conclusion that the scale in question does not fit the data, that the theoretical assumptions are not appropriate for that particular situation. Under these circumstances, no scale is obtained. But when the empirical fit of data to the theory is acceptable, the result is a scale whose properties are based to some extent on empirical law rather than arbitrary definition alone. It is with such fundamental measurement procedures in scaling that this article is concerned.

**Two broad classes of scaling models.** The fundamental scaling procedures can be grouped into two broad classes: those concerned with the measurement of stimulus attributes and those concerned primarily with measurement of individual difference or responder attributes. Different models within each class represent alternative rationales for attaining essentially the same objective. The objective itself ordinarily differs between classes.

### Stimulus models

The stimulus models use responses of subjects to stimuli to establish scales for attributes of stimuli. Scale values are assigned only to the stimuli. The subject, or group of subjects, is characterized by the entire scale. Individual or group differences in this context refer to *differences between scales*, and not to differences in scale values on the attribute. The stimulus models are therefore concerned with psychological attributes of stimuli for subjects or groups of subjects.

In the stimulus model, the subject himself acts as a judge or evaluator of the stimuli. His task is a particularized one—he does not simply judge or



respond to the stimuli *per se* but, rather, evaluates the stimuli with respect to some specified attribute. Hence, the same set of stimuli can be scaled with respect to several different attributes. A set of statements all designed to reflect an attitude toward some topic might, for example, be scaled with respect to such attributes as degree of favorability toward the topic, relevance, complexity, or intensity. The subject's task in each case is to evaluate the stimuli solely with respect to the attribute of interest. The resulting scale is then indicative of how these particular statements are perceived with respect to the defined attribute by that particular subject or group of subjects.

Stimulus models are used to obtain both psychophysical and psychometric scales. The models appropriate for psychometrics are actually a subset of the psychophysical models, but the measurement problem itself differs considerably in the two fields. Measurement in psychophysics has been concerned primarily with determining functional relations between physical variables and the corresponding subjective attributes. The usual experimental strategy holds constant or controls all relevant physical dimensions but one, and then determines how the subjective attribute varies as a function of that single physical variable. The resulting psychophysical magnitude function is thus a relation between two variables that is expected to hold when other physical variables influencing the subjective attribute are fixed or controlled as specified. For example, a psychophysical function for loudness might relate the perceived loudness of 1,000-cycle sinusoidal tones of a given duration to sound energy measured at the ear. Other functions ordinarily are needed for white noise, for clicks, or for tones of another duration or a different frequency. Although each separate function is monotonic for the relevant physical variable, no such simple relation exists relating sound energy to subjective loudness for stimuli that vary on several relevant dimensions. Development of equations specifying the loudness of any stimulus that possesses sound energy would be a great deal more complex. [See HEARING; PSYCHOPHYSICS.]

In psychophysics the physical variable determines the set of stimuli to which the results apply. In psychometrics, it is more nearly the reverse. Psychometric scaling begins with the psychological attribute. The procedures are applicable to any set of stimuli possessing the attribute, and the problem is to scale the set of stimuli as it exists. The stimulus set is typically complex, varying with respect to other subjective dimensions and with respect to a host of correlated physical variables. There is, of

course, no single, monotonically related physical correlate. Monotone relations are to be expected only when all but one of the physical correlates are held constant. Complexity of the stimulus set is therefore the chief characteristic distinguishing psychophysical from psychometric scaling problems. Several major consequences of this complexity are listed below. [See PSYCHOMETRICS.]

*Recognizability of individual stimuli*—The subject can often identify psychometric stimuli because of their variation on extraneous attributes. Since he may tend to remember earlier responses to the same stimulus, independent judgments on successive trials cannot routinely be expected. Replication over trials with the same subject is therefore troublesome. Most psychometric scales have been based on groups of subjects, so that the recognition problem is not so severe.

*Ordinal properties*—When there is awareness of a monotonically related physical variable, knowledge of order on the subjective attribute is automatically implied. For the complex stimulus sets of psychometrics, order of the stimuli is typically an unknown that must be determined. Experimental procedures that require prior knowledge of order of the stimuli often then cannot be used.

*Production techniques*—When a physical variable is monotonically related to a psychological attribute, there is ordinarily a procedure for easily generating a virtually continuous set of stimuli that are ordered with respect to the psychological attribute. This in turn makes possible the use of the various production techniques—procedures that require the subject to adjust a variable, i.e., to generate a new stimulus, with specified properties. On the other hand, with psychometric stimuli such production techniques typically cannot be used: the subject cannot on demand paint a painting halfway between two standards in aesthetic quality, write an essay twice as good as a standard, or create a worker whose merit falls between that of two others.

*Interpolation*—A monotonic physical attribute can be used as a basis for interpolation. The stimulus that could be judged greater than a standard 75 per cent of the time can be estimated from data obtained by using other stimuli. In addition, a subset of stimuli can be used to establish the functional relation between the two variables. Given the relation, the scale values of all remaining stimuli can be determined numerically or graphically, without resort to additional experimental observations. Neither technique is available for the complex stimulus sets of psychometrics, the closest approximation of which is the establishment of a scale of

standards against which new stimuli can be evaluated by direct comparison.

The stimulus scaling models can be classified into two types on the basis of the rationale used to obtain scale properties above and beyond order. In the *variability* models, the discriminability of the stimuli, or the variability of judgment with respect to the stimuli, is used as the basis for deriving the higher-level metric scale properties. In the *quantitative judgment* models, the additional metric properties are obtained directly from the quantitative aspects of the judgments themselves.

**Variability models.** Perhaps the best-known variability model is that developed by Thurstone (1927). Thurstone's general model begins with a set of elements called *discriminal processes*. Each element has a value on a postulated, underlying psychological continuum. Presentation of a stimulus to a subject results in arousal of a discriminational process. Different presentations of the same stimulus may arouse different discriminational processes—there is thus no one-to-one relation between the two. Rather, a distribution of discriminational processes is associated with each stimulus. Presentation of a stimulus is equivalent to sampling a discriminational process from the distribution associated with that stimulus. [See CONCEPT FORMATION; LEARNING, article on DISCRIMINATION LEARNING.]

In Thurstone's model it is postulated that these distributions (one for each stimulus) are normal with means  $s_i$  and variances  $\sigma_i^2$ . The mean is taken as the scale value of the stimulus. The standard deviation,  $\sigma_i$ , called the *discriminational dispersion*, is the index of the ambiguity or confusability of the stimulus; the smaller the  $\sigma_i$ , the less ambiguous and the more readily discriminable the stimulus. The general model, therefore, provides both location and spread parameters for the stimuli. The conceptual framework of the model is illustrated in Figure 1.

We will consider here two particularizations of

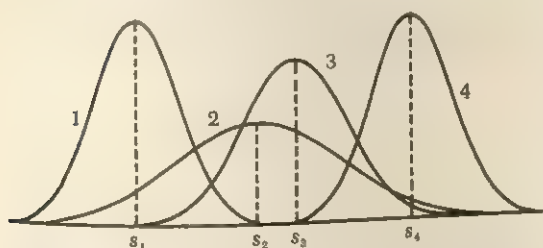


Figure 1 — Distributions on the psychological continuum of discriminational processes associated with four stimuli

Source: From Torgerson 1958, p. 157.

the Thurstone model, one developed for comparative judgments and another developed for categorical judgments.

**The law of comparative judgment.** The law of comparative judgment postulates that in any comparison of two stimuli, the psychological difference between the stimuli is indirectly measured by the relative frequency with which the difference is perceived. For example, two stimuli,  $j$  and  $k$ , are presented to the subject, with instructions to indicate which one appears greater with respect to some designated attribute. Since each stimulus is held to arouse a discriminational process,  $d_j$  or  $d_k$ , the subject selects as greater the stimulus which on that occasion arouses the discriminational process having the greater value on the underlying continuum.

If the difference  $d_k - d_j > 0$ , stimulus  $k$  is selected; if  $d_k - d_j < 0$ , stimulus  $j$  is selected. Exact equality is not allowed. Each paired presentation is formally equivalent to drawing an element from a bivariate normal distribution. An indefinitely large number of such observations generates a distribution of *discriminational differences*. From a large number of such judgments one can also determine the proportion of times that stimulus  $k$  is judged greater than stimulus  $j$ .

One of the characteristics of the normal distribution that makes it particularly valuable for distribution models of this type is that differences between paired observations drawn from a bivariate normal distribution are normally distributed. The mean of the difference distribution is equal to the difference between the original means, or scale values,  $s_k - s_j$ ; the variance is  $\sigma_j^2 + \sigma_k^2 - 2r_{jk}\sigma_j\sigma_k$ , where  $r_{jk}$  represents the correlation between the pairs of discriminational processes. If we define  $p_{jk}$  as the probability that stimulus  $k$  is judged greater than stimulus  $j$  (that is, that  $d_k - d_j > 0$ ) on any given presentation, and define  $x_{jk}$  as the unit normal deviate corresponding to  $p_{jk}$ , then the general form of the law of comparative judgment for  $n$  stimuli is given by the set of equations

$$(1) \quad s_k - s_j = x_{jk} \sqrt{\sigma_j^2 + \sigma_k^2 - 2r_{jk}\sigma_j\sigma_k}, \\ j, k = 1, 2, \dots, n.$$

The quantity  $x_{jk}$  is the (theoretical) observable; the remaining terms are unknowns. Given  $n$  stimuli, there are  $n(n-1)/2$  possible equations and, since both the origin and unit are arbitrary, there are  $2(n-1) + n(n-1)/2$  unknowns. Hence there are always more unknowns than independent equations, and further restrictions on the data are necessary to establish scale values. Three such specializations have been proposed, of which only one, Thurstone's case v, the most restrictive, has been



found generally useful. In case v, it is assumed that the value under the radical in (1) is constant, i.e., that the distributions of discriminial differences all have the same variance. Then, arbitrarily defining the variance as unity, for convenience, there remain simply the  $n(n-1)/2$  equations

$$(2) \quad s_k - s_j = x_{jk}, \quad j, k = 1, 2, \dots, n.$$

Efficient procedures for estimating the scale values from these equations are available. The scale obtained is an interval scale, because the unit of measurement must be defined arbitrarily, and a constant can be added to all of the scale values without changing the solution. Furthermore, the interval properties are based on empirical law. Consider any three ordered stimuli,  $i < j < k$ , and the three unit normal deviates  $x_{ij}$ ,  $x_{jk}$ , and  $x_{ik}$ . From (2), since the numerical differences between scale values must add,

$$(s_k - s_j) + (s_j - s_i) = (s_k - s_i), \quad \text{for all } i < j < k,$$

it is clear that the model requires that

$$(3) \quad x_{jk} + x_{ij} = x_{ik}, \quad \text{for all } i < j < k.$$

The additivity-of-difference rule can therefore be said to provide the empirical basis for the interval properties of the final scale. Actually, of course, one would not expect (3) to hold exactly for empirical data. Among other things, the obtained proportions on which the deviates are based are only estimates and not population values. A statistical test for the goodness of fit of the model to data has been provided by Mosteller (1951). Graphical and numerical procedures can also be used to aid the investigator in deciding whether the fit is close enough for his purposes even though it does not meet the statistical criterion. Given estimates of the scale values of the stimuli, one can compute the set of theoretical proportions that would fit the estimated scale values perfectly. The discrepancies between these derived proportions and the corresponding observed proportions can then be examined. Alternatively, for any two stimuli,  $i$  and  $j$ , one can plot the obtained values of  $x_{ik}$  versus  $x_{jk}$  for all values of  $k$ . From (2) it is clear that the model requires that the plotted points fall along a straight line with unit slope. [See GOODNESS OF FIT.]

*The law of categorical judgment.* Several additional postulates are used to adapt the general Thurstone model to categorical data and to derive a set of equations relating stimulus values and category boundaries to the relative frequency with which each stimulus is judged to be in each category. In the usual experimental procedures, the

subject is required to rate or sort a set of  $n$  stimuli into a fixed number of ordered categories. Graphic rating procedures or arrangement methods can also be used, but these require that the data first be artificially divided into a limited set of ordered categories.

Thurstone's general model illustrated in Figure 1 deals with a set of  $n$  stimuli and their associated distributions of discriminial processes, with means  $s_j$  and standard deviations  $\sigma_j$ . The present extension adds parallel notions dealing with the location of boundaries between adjacent categories.

Data are presented in the form of ratings of stimuli on an  $m+1$  step scale. It is assumed that the psychological continuum is divided by the  $m$  category boundaries into  $m+1$  intervals, which may or may not be equal in extent. In the original version of the model, the boundaries were treated as fixed. We shall follow the more general version that allows the boundary locations to vary from observation to observation. The location of the  $g$ th category boundary is treated as a random variable with mean  $t_g$  and variance  $\sigma_g^2$ . If we assume that these distribution functions are also jointly normal, then precisely the same reasoning that led to (1) leads to the set of  $mn$  equations for the general law of categorical judgment:

$$(4) \quad t_g - s_j = x_{jg} \sqrt{\sigma_j^2 + \sigma_g^2 - 2r_{jg}\sigma_j\sigma_g}, \\ j = 1, 2, \dots, n, \\ g = 1, 2, \dots, m,$$

where  $t_g - s_j$  denotes the difference between the mean location of category boundary  $g$  and the mean location of stimulus  $j$ . Eq. (4) is again too general to be of more than academic interest. As was true for (1), there are always more unknowns than equations. One common restricted version of the general equation is based on the assumption that the radical is a monotone function of the stimulus standard deviations,  $\sigma_j$ . This would occur, for example, if both the correlation term and the variances of the category boundaries were constant, or if constant variances  $\sigma_g^2$  and independence ( $r_{jg} = 0$ ) were assumed. If we now let

$$\alpha_j^2 = \sigma_j^2 + \sigma_g^2 - 2r_{jg}\sigma_j\sigma_g,$$

we can then write the  $nm$  equations

$$(5) \quad t_g - s_j = x_{jg}\alpha_j, \quad j = 1, 2, \dots, n, \\ g = 1, 2, \dots, m.$$

Eqs. (5) are formally identical to the equations underlying the "method of successive intervals" (Saffir 1937; Gulliksen 1954), although in the traditional version the assumption of fixed boundaries leads to the interpretation of  $\alpha_j$  as the standard

deviation (discriminal dispersion) of stimulus  $j$  rather than as a more general spread parameter only monotonically related to  $\sigma_j$ .

Efficient and practical estimation procedures have been developed for both complete and incomplete data matrices and are discussed in Torgerson (1958).

Another common restricted case assumes that the difference distributions all have the same variance. Then, if the unit of measurement is defined so that the value under the radical in (4) is 1, we can write the  $nm$  equations,

$$(6) \quad t_{ij} - s_j = x_{ij}, \quad \begin{matrix} j = 1, 2, \dots, n, \\ g = 1, 2, \dots, m. \end{matrix}$$

Eqs. (6) represent the most restricted version of the law and underlie Attneave's method of graded dichotomies (1949), Garner and Hake's equidiscriminability scaling procedure (1951), and Edwards and Thurstone's method of successive intervals (1952). Note that this version allows a single parameter for both stimuli and category boundaries, whereas the other version allowed a second parameter for the stimuli. The final scale, if achieved, is therefore an equidiscriminability or equiconfusability scale, with properties similar to those of (2) for comparative judgments.

*The logistic model for choice data.* The logistic model is the major alternative to Thurstone's general model for obtaining a scale based on discriminability or variability of judgment. The model was developed for the paired comparison case by Bradley and Terry (1952). Luce's choice axiom (1959) extended the model to choice data in general and gave the procedure a more elegant, axiomatic basis.

Luce's version of the model requires, in essence, that the subjects' choices behave in a manner analogous to conditional probabilities. Where  $S$  and  $T$  are both subsets of a set  $U$  and where  $p(T)$ , the probability of choosing an element of  $T$ , is greater than 0, then the formula for the conditional probability of  $S$  given  $T$ ,  $p(S|T)$ , is  $p(S \cap T)/p(T)$ . The interesting property of conditional probability for present purposes is the multiplication rule. Given three subsets, such that  $R \subset S \subset T$ ,

$$p(R|S) \cdot p(S|T) = p(R|T).$$

Luce's axiom, for the case where the probability of choosing any element is neither 0 nor 1, is the empirical analogue:

$$(7) \quad p_S(R) \cdot p_T(S) = p_T(R).$$

That is, the probability that the subject will choose an  $R$ -element when presented with the set  $T$  is equal to the probability that he will choose an  $R$ -

element when presented with the set  $S$  multiplied by the probability that he will choose an  $S$ -element when presented with the set  $T$ .

An alternative form of the axiom is known as the constant ratio rule (Clarke 1957). Let  $x$  and  $y$  denote any two stimuli, let  $p(x, y)$  denote the probability that stimulus  $x$  is chosen when only  $x$  and  $y$  are presented, and let  $p_S(x)$  denote the probability that stimulus  $x$  is chosen when the subset  $S$  is presented. The constant ratio rule states that

$$(8) \quad \frac{p(x, y)}{p(y, x)} = \frac{p_S(x)}{p_S(y)} = C_{xy},$$

where  $C_{xy}$  denotes the constant ratio of the probability of choosing stimulus  $x$  to the probability of choosing stimulus  $y$ . The rule requires that for any set  $S$  containing both stimulus  $x$  and stimulus  $y$ , the ratio of the probability of choosing stimulus  $x$  to that of choosing stimulus  $y$  be invariant. For a given pair of stimuli the ratio is constant regardless of either what or how many other stimuli are included in the set from which the choice is made. The complete experiment for evaluating Luce's axiom or the constant ratio rule would therefore require the experimenter to obtain separate estimates of the probability of choosing stimulus  $x$  (for all  $x \in T$ ) for every subset of  $T$  containing  $x$  as a member.

When the requirements of the axiom are met, then there exists a  $v$ -scale such that for every subset of stimuli  $S$ ,

$$(9) \quad p_S(x) = \frac{v(x)}{\sum_{y \in S} v(y)},$$

where  $v(x)$  denotes the scale value of stimulus  $x$ . From (9) and the constant ratio rule it is also clear that

$$(10) \quad \frac{v(x)}{v(y)} = \frac{p_S(x)}{p_S(y)} = \frac{p(x, y)}{p(y, x)}.$$

Given the axiom and the definition, the scale values are determined up to multiplication by a positive constant, i.e., the  $v$ -scale is a ratio scale. It should be noted, however, that the ratio property depends on the particular definition chosen, as well as on the requirements placed on the data by the axiom. The empirical requirements of the choice axiom may limit the scale to determination up to a multiplicative constant and an exponent. Eq. (9), in effect, specifies a particular exponent in the same way that the unit was specified in (2).

The axiom places two rather severe restrictions on the data. Consider first the paired comparison case, and let  $x, y, z$  denote any three stimuli. Then



the probability ratios must follow a multiplication rule:

$$(11) \quad \frac{p(x,z)}{p(z,x)} = \frac{p(x,y)}{p(y,x)} \cdot \frac{p(y,z)}{p(z,y)}.$$

Hence, any two probabilities determine the third. It can also be shown that the pairwise probabilities determine all probabilities of choice from larger sets and, in fact,

$$(12) \quad p_s(x) = \frac{1}{\sum_{y \in S} \frac{p(y,x)}{p(x,y)}}.$$

An adequate test of the model as stated would require evaluation of the extent to which both restrictions are met.

The simpler version of the logistic model, as developed by Bradley and Terry, is limited to paired comparison data and, hence, only the first restriction is applicable. Efficient estimating procedures and tests of goodness of fit are available (Bradley & Terry 1952; Bradley 1954).

*Thurstone's case V and the logistic model.* If we consider only the paired comparison situation, then the equation relating scale values to observable proportions for the logistic model can be written simply as

$$(13) \quad p(x,y) = \frac{v(x)}{v(x) + v(y)}$$

or, equivalently, as

$$(14) \quad \frac{v(x)}{v(y)} = \frac{p(x,y)}{p(y,x)}.$$

Eq. (14) shows that in the  $v$ -scale a simple function of the observable is interpreted as a ratio of scale values. From (11) it is apparent that the metric properties of the  $v$ -scale are based on a multiplication-of-ratios rule, with log interval properties. In Thurstone's case  $v$ , the corresponding basis for the metric properties was an additivity-of-differences rule, which yielded an interval scale.

Bradley and Terry, as well as Luce, have shown that the logistic model can also be used to generate an interval scale based on additivity of differences. If one defines  $r(x) = \ln v(x)$  and  $w(x,y) = r(x) - r(y)$ , then (14) can be written

$$(15) \quad \ln \frac{p(x,y)}{p(y,x)} = r(x) - r(y) = w(x,y),$$

and from (13),

$$(16) \quad p(x,y) = \frac{1}{1 + e^{-w(x,y)}},$$

which is a simple form of the logistic function.

The logistic function and the normal ogive are markedly similar over most of the empirically in-

teresting range. As a result, a close fit of the data to Thurstone's case  $v$  insures a fit to the logistic model that can be only trivially better or worse. For all practical purposes, the values from the  $r$ -scale of the logistic model will be linearly related to the scale values from the Thurstone case  $v$  model. And both, of course, will be logarithmically related to the  $v$ -scale values.

It might be noted that the Thurstone model can also be easily modified to yield a log interval scale similar to the  $v$ -scale of the logistic model. We need only to substitute log normal distributions for normal distributions and subjective ratios for subjective differences. The new scale will be exponentially related to the Thurstone case  $v$  scale, and will therefore be related to the  $v$ -scale by a power transformation. In general, any model in which the observable is interpreted as a subjective difference in scale values can be paralleled by one in which the observable is interpreted as a subjective ratio. The difference-based scale will be logarithmically related to the ratio-based scale, and the fit of one model to any given set of data will be precisely matched by the fit of the other, parallel model.

*Quantitative judgment models.* The variability methods require only ordinal judgments by the subject; the higher-level, metric properties of the obtained scales are derived through models relating variability of judgment to subjective differences or ratios. In the quantitative judgment models, the higher-level properties are obtained *directly* from the subject's judgments. There are two classes of quantitative judgment methods: the difference, or distance, methods and the ratio methods. The difference methods have long been used to develop scales in the social sciences (Thurstone & Chave 1929). Serious application of the ratio methods in the context of social science is more recent (see, for example, the summary of research in Stevens 1966). In the difference methods, the metric properties of the obtained scale depend upon the ability of the subject to evaluate or equate subjective differences. In the ratio methods, the metric properties depend upon his ability to evaluate or equate subjective ratios. Scales obtained from the two types of procedures are not ordinarily linearly related.

*Difference methods.* Difference methods include the methods of equisection, equal-appearing intervals, category rating, and category production. The methods differ from one another in the specific nature of the experimental procedure employed, but they all have the same objective: to obtain a scale in which the differences in numbers assigned to the stimuli directly reflect the corresponding subjective differences between the stimuli.

In equisection, the subject produces a set of stimuli that divides a given range of the attribute into subjectively equal steps. In equal-appearing intervals and category rating he sorts or rates the stimuli into a given set of subjectively equally spaced steps or categories. In category production, he produces on each trial a stimulus that corresponds to one of a set of equally spaced categories.

The usual procedures do not include built-in checks for the adequacy of the assumption that the subject can produce such judgments in an internally consistent manner. The assumption does place empirical restrictions on the behavior of the judgments, however, so appropriate checks are possible. The simplest and most direct check uses the method of bisection. Here, the subject first bisects the interval between two standard stimuli,  $a$  and  $e$ , to produce a new stimulus,  $c$ , halfway between. He next bisects the interval  $a-c$ , producing a stimulus  $b$ , and bisects the interval  $c-e$ , producing a stimulus  $d$ . Finally, he bisects the interval between  $b$  and  $d$ . If the subject's bisections are made in an internally consistent manner, the stimulus produced by the final bisection between  $b$  and  $d$  will be equal to  $c$ , the stimulus produced by the bisection of  $a$  and  $e$ .

For psychometric stimuli, where production techniques are unavailable, a more or less equivalent procedure involves imbedding a given subset of stimuli in several different larger sets. If the subject's rating or sorting judgments fulfill the requirements of an interval scale, the scale values assigned to the members of the subset will be linearly related across the several experiments.

*Ratio methods.* Ratio methods include the methods of fractionation, constant sum, magnitude and ratio estimation, and magnitude and ratio production. In all of these methods, the subject's task is to adjust stimuli or assign numbers so that the ratios between the numbers correspond to the subjective ratios between the stimuli. In fractionation, the subject adjusts a variable to be equal to  $1/k$  of a standard. In the constant sum method, he divides 100 points between two standards so that the ratio between the assigned points corresponds to the subjective ratio. In the magnitude estimation procedure, his task is to assign numbers so they are proportional to the subjective magnitudes, whereas in the ratio estimation method, he assigns numerical ratios to pairs of stimuli. In the corresponding production procedures, the roles of numbers and stimuli are reversed: the experimenter chooses the numbers and the subject produces the appropriate stimuli or stimulus ratios.

Only the constant sum method has a built-in

check for adequacy of the basic assumption. But again, appropriate checks of internal consistency are available for the other methods and, in fact, parallel those for the difference methods.

*Differences versus ratios.* The logarithmic, or nearly logarithmic, empirical relationship between difference scales and ratio scales has led to the conjecture that the subject's quantitative judgments are based on a single perceived quantitative relation (Torgerson 1961). According to this argument, when the subject is told to equate differences, he interprets the subjective relation as a difference. When he is told to equate ratios, he interprets the same relation as a ratio. The conjecture receives additional support from the behavior of quantitative judgments when the direction of the attribute is reversed, i.e., when the subject reports on smoothness rather than roughness, or darkness rather than lightness. For difference scales, the attribute and its reverse are linearly, though negatively, related. Stimuli that are separated by equal differences in one direction remain equally different when the direction of judgment is reversed. For the ratio scales, the attribute and its reverse are reciprocally related. Here the equal ratio property is invariant, but differences of course change markedly.

The empirical results based on quantitative judgment are thus similar to the formal limitations of the variability methods. Each set of procedures yields scales on which either differences or ratios, but not both, are empirically meaningful. Additional procedures seem necessary to obtain a scale where both relations have an empirical basis. One such procedure has been devised by Cliff (1959).

### Individual difference models

The individual difference, or responder, models use subjects' responses to stimuli as a way of establishing scales for attributes on which the subjects (responders) themselves differ. The models are based upon individual differences; an ultimate aim is to assign scale values to the subjects to represent amount or degree of the attribute possessed by each. Stimuli may also be assigned scale values on a parallel attribute, or they may be assigned parameter values relating them in some more general manner to the attribute of interest. Individual difference models are therefore those in which a response to a stimulus is interpreted as a function of both the scale value of the responder and the parameter values of the stimulus.

The relation between the task set for the subject and the attribute continuum differs markedly between the stimulus models and the individual dif-



ference models. In the stimulus models, the task is a particularized one—the subject evaluates stimuli with respect to the attribute being measured. In the individual difference models, the task itself has no obvious relation to the underlying attribute. In the *categorical response* methods, the subject simply indicates whether or not (or the extent to which) he agrees with, likes, or can pass a particular item. In the *comparative response* methods, he indicates which item he agrees with most, or which he prefers. In neither case, however, is the attribute of interest one of agreement or preference. It is, rather, a latent or underlying attribute that serves to explain or account for the variation in the responses of the subjects to the stimulus items.

A concrete example might serve to clarify the distinction. Suppose we are given a set of samples of coffee that differ *only* in the amount of sugar that has been added. The samples are presented, two at a time, to a population of subjects. Each subject chooses the member of each pair that he prefers. If a *stimulus* model were successfully applied to the data, the result would be a scale of *preferability* of the samples, ranging from the least preferred to the most preferred for that population of subjects. But if an *individual differences* model were successfully applied to the data, the result would be a scale of *sweetness*, with the stimuli ranging from least sweet to most sweet and *each subject characterized by a scale value denoting the degree of sweetness that he considers ideal*. The subjects' responses do not tell us, of course, that the underlying variable is degree of sweetness. If the coffee samples had varied in strength or in amount of cream added rather than in sweetness, their responses would still have been of exactly the same form. In general, the individual differences models are explicitly concerned solely with the question of whether or not the responses of a given group of subjects to a given set of stimuli can be accounted for by a single underlying attribute. They do not tell one how to select the stimuli in the first place or, given that the stimuli form a scale, how the underlying attribute should be interpreted.

Individual differences models can be classified in a number of ways. One important difference, which will be used to organize the remainder of this section, is the distinction between the simple *deterministic* models and the more elaborate *probabilistic* models. Deterministic models are developed in terms of the ideal case; there is no provision in the model itself for unsystematic variance. Since the ideal case virtually never occurs, the empirical question is one of the extent to which the ideal approximates the actual behavior of the subjects.

The probabilistic models represent more realistic attempts to account for the behavior of the subjects, provide explicitly for unsystematic variance, and, at least theoretically, lend themselves to statistical evaluation of goodness of fit.

**Simple deterministic models.** Three general models will be considered here: Coombs's distance models for preferential choice data, Guttman's scalogram model for categorical responses to monotone items, and the scalogram models for categorical responses to point items.

**Distance models.** In Coombs's distance models, both subjects and stimuli are represented as points along an underlying, latent continuum. According to the general model, the subject's preferential choices among stimuli are determined by the relative distances of the stimuli from that subject. A stimulus located at the same point as the subject represents his ideal, the one he would prefer above all others. As distance from the subject increases, the desirability of the stimulus to that subject decreases. Subject A thus will prefer stimulus X to stimulus Y when the distance  $d_{AX}$  is less than the distance  $d_{AY}$  on the underlying continuum.

Any of a large number of experimental procedures might be used to obtain a preferential ranking of the stimuli—called a *qualitative I-scale*—for each subject. The analytical problem is to determine whether the set of I-scales could have been derived from distances between subjects and stimuli on a common underlying attribute, and, if so, to construct the joint scale—or *J-scale*—giving the required locations of stimuli and subjects. Coombs's *unfolding technique* (1950; 1964) is a procedure for accomplishing this goal. The end result is a scale which orders the stimuli, orders the midpoints between pairs of stimuli, and locates the subjects in segments of the scale defined by the interval between adjacent stimulus midpoints.

The order of the midpoints provides information on the relative size of the intervals between some, but not all, of the stimuli. For example, given four stimuli in order ABCD, the midpoint between B and C precedes that between A and D, when the distance  $d_{BC} > d_{AD}$ . Scales such as this, which provide not only an order for the elements themselves but also an order for some of the distances between elements, are called ordered metric scales.

With suitable modifications (Coombs 1953), the general approach described above can also be applied to data obtained from the "order  $k|n$ " and "pick  $k|n$ " methods. In the "order  $k|n$ " methods, the subject rank-orders only the first  $k$  out of the  $n$  stimuli presented to him, and in the "pick  $k|n$ " methods, he simply chooses the  $k$  out of  $n$  stimuli

he prefers most. Procedures for ordering the stimuli and for locating the subjects in ordered classes are given in Coombs (1953) and Torgerson (1958).

*Monotone and point scalogram models.* The scalogram models for categorical data are concerned with subjects and items, where an item is defined as any stimulus or procedure that partitions the set of subjects into two or more mutually exclusive categories. In these models, both subjects and category boundaries can be considered as points on an underlying continuum. The scaling problem is to determine whether the subjects and category boundaries can be positioned along the continuum so that all subjects are located within the appropriate categories of all items.

Three types of ideal items are of interest: the monotone multicategory item, the monotone dichotomous item, and the nonmonotone, or point, item.

A *monotone item* divides the underlying continuum into as many segments as there are response categories, with each category corresponding to one and only one segment. The following multicategory item provides an example of the general form:

My weight is (a) between 150 and 175 pounds;  
(b) 150 pounds or less; (c) 175 pounds or more.

*Dichotomous monotone items* provide only two response alternatives. For example,

I weigh over 150 pounds.      True \_\_\_\_\_  
False \_\_\_\_\_.

Different items ordinarily partition the underlying continuum in different places. In general,  $n$  monotone items, with a total of  $m$  categories in all, partition the underlying continuum into  $m - n + 1$  segments.

Guttman's general scalogram procedure (see Stouffer et al. 1950) is a routine for determining whether the  $n$  items (a) are monotonic and (b) can be considered as partitioning the same underlying attribute for the given population of subjects. If so, the final result is, first, a rank order of the item category boundaries on the underlying continuum and, second, the locations of subjects in the segments defined by adjacent category boundaries. The general procedure can be used with any combination of multicategory and dichotomous monotone items. Somewhat simpler procedures are available if all items are dichotomous.

*Nonmonotone, or point, items* also partition the underlying continuum into segments, but not in a one-to-one fashion. An example of one form of nonmonotone item is

My weight is between 150 and 175 pounds.  
True \_\_\_\_\_  
False \_\_\_\_\_.

The positive response to this type of nonmonotone item corresponds to a single segment—in this example, the range from 150 to 175 pounds. But a negative response corresponds to two segments: subjects weighing over 175 pounds and those weighing under 150 pounds both respond negatively.

In order to evaluate the scale hypothesis and to obtain an ordinal scale, it is necessary for the segments corresponding to the positive alternatives to overlap. If no overlap occurs, the data contain information sufficient only to locate the subjects into a set of unordered categories. The procedures for uncovering the underlying scale for nonmonotone items are different from, but conceptually similar to, those appropriate for monotone items.

An alternative ideal model for nonmonotone items treats the positive category of the item as a point and represents each subject by a segment of the underlying continuum. An example of an item which might fit such a model better would be

I weigh about 150 pounds.      Yes \_\_\_\_\_  
No \_\_\_\_\_.

Here, subjects might differ with respect to their ideas of the range encompassed by the term "about." Since the two models are formally equivalent—they differ only in the interchange of the role of subject and item—the same analytical routines are applicable.

*Probabilistic models: the general theory.* Most of the models developed for measuring individual difference attributes can be considered within the framework of a single general probabilistic model whose conceptual basis is essentially that of Lazarsfeld's latent structure analysis (1954; see also Stouffer et al. 1950), in which the attribute to be measured is represented formally by an underlying or latent space of one or more dimensions, subjects are represented by points in the latent space, and the stimuli are treated as variables. Stimuli may be psychological test items, attitude statements, rating scales, or, in general, anything that classifies the subjects into two or more exclusive and exhaustive categories. The observations, or *manifest data*, thus consist of an  $N$  by  $n$  matrix of the scores of the  $N$  subjects on the  $n$  stimuli. The stimulus variables are characterized in the general model by the parameters of the equation describing the regression of that variable on the underlying latent attribute.

An important general notion is the principle of local independence. Since each stimulus variable is related to the underlying attribute, the stimulus variables will be related to each other in some fashion. The principle of local independence re-



quires that the systematic relations among the stimuli be accounted for entirely by the separate relations of each stimulus variable to the underlying attribute. Thus, at any given point in the latent space, the stimuli are all mutually independent.

The over-all theory or conceptual framework clearly is completely general. In order to turn it into a testable model—one which places restrictions on the data—additional postulates or conditions are necessary. Three important considerations are the nature of the manifest data required, the nature of the postulated latent attribute, and the form of the regression equation relating the stimulus variables to the latent attribute. A particular model or testable specialization of the general theory depends on making a decision about each of the three. [See LATENT STRUCTURE.]

Three levels or types of manifest response data are of major concern here, but obviously others are possible: *Dichotomous items*—the stimulus variable classifies the subject into one of two categories; *ordered classes*—the stimulus variable classifies the subject into one of several ordered categories; *interval scale*—the stimulus variable provides an interval scale value for the subjects.

Three restrictive assumptions that can be placed on the postulated latent attribute are the following: the subjects are assumed to be located only at a relatively small number of points or regions in the space; the underlying attribute is assumed to be unidimensional; and the underlying attribute is assumed to be a multidimensional Euclidean space.

Four types of regression equations that relate the stimulus variables to the latent attribute are the various step functions, the linear equation, the normal ogive equation, and the logistic equation.

The models also differ in whether or not they provide for or allow variance about the regression line or surface. The distinction here is essentially that made earlier between deterministic and probabilistic models. The deterministic models for categorical data, which make no provision for extraneous variance, can be considered as limiting cases for certain of the probabilistic versions.

Some specializations of the general theory. All of the models described in this section can be treated as specialized instances of general regression theory. In all cases, subjects can be considered as points in an underlying attribute space, stimuli as manifest variables related to the underlying space by a regression equation, and interrelations between manifest variables in terms of their separate relations to the underlying attribute space. The models differ in the kind of data required, the kind of space allowed, the form of the postulated regression equation, and in whether or not explicit provision is made for variance about the postulated regression surface. Table 1 provides a summary of the characteristics of and differences between a number of the more common models.

*Lazarsfeld's latent class model.* In Lazarsfeld's latent class model the stimuli (items) are dichotomous and are scored either as 0 or as 1. The latent space is nominal—subjects are located at discrete

Table 1 — Characteristics of some individual difference models that fall within the general regression approach

	DETERMINISTIC			PROBABILISTIC							
	General scalogram	Dichoto- mous scalogram	Multidi- mensional extensions of scalogram	Latent class	Latent profile	Latent distance	Linear	Two- factor	Multiple factor analysis	Normal ogive	Logistic
Observable											
Dichotomous		X	X	X			X			X	X
Ordered classes	X										
Interval					X			X	X		
Latent space											
Points				X	X						
Unidimensional	X	X									
Multidimensional			X			X	X	X		X	X
Regression									X		
Step function	X	X	X	X	X	X					
Linear								X	X	X	
Normal ogive											
Logistic										X	
Variance about regression											
Yes				X	X	X	X	X	X	X	X
No	X	X	X								

points or, equivalently, along arbitrarily ordered intervals of a unidimensional continuum. The regression equation is then a step function, and provision is made for variance about the regression line.

If the classes are ordered arbitrarily, the height of the regression line for an item  $j$ , over the interval defined by class  $s$ , is the average score of the subjects in class  $s$  on that item. Since only scores of 0 and 1 are used, the height of the regression line for class  $s$  is also the probability that a person in class  $s$  will answer item  $j$  correctly. This probability is denoted by  $v_{js}$ . The unknowns to be determined are the set of probabilities  $v_{js}$  and the proportion of subjects in each class,  $\pi_s$ .

The observational information can be expressed by the following set of manifest proportions:

$p_j$  = the proportion of subjects in the entire sample answering item  $j$  positively;

$p_{jk}$  = the proportion of subjects answering both item  $j$  and item  $k$  positively;

$p_{jkl}$  = the proportion answering items  $j$ ,  $k$ , and  $l$  positively;

$\vdots$   $\vdots$

Lazarsfeld has shown that this information can be incorporated into a set of accounting equations, and several procedures have been devised for solving such equations (Green 1951; Anderson 1954). The solutions themselves do not directly indicate the class membership of individual subjects but, rather, give only the proportion of subjects in a class. Additional procedures are needed to assign subjects to classes.

**Gibson's latent profile model.** Gibson's latent profile model (1959), which differs from Lazarsfeld's only in the type of data required, is designed for manifest responses measured on an interval scale. Since the equations for latent profile analysis are formally identical to those of latent class analysis, no new analytical procedures are needed for their solution.

**Lazarsfeld's continuous models.** In Lazarsfeld's continuous models the stimuli are dichotomous items scored either as 0 or as 1, but the latent space is generally assumed to be a unidimensional continuum, although multidimensional extensions have also been suggested. Hence, the several continuous models proposed by Lazarsfeld differ from each other only in the form of the regression line. Where  $x$  = the latent variable,  $f_j(x)$  = the regression of item  $j$  on  $x$ , and  $g(x)$  = the density function of

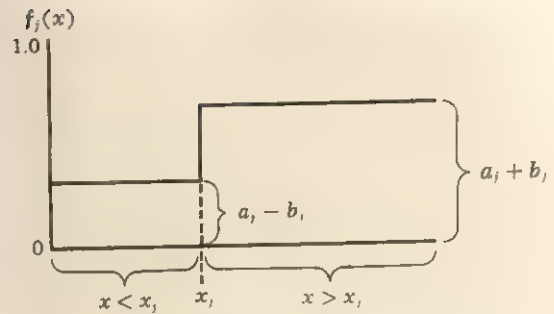


Figure 2 — Form of item regression or trace line for the latent distance model

Source: From Torgerson 1958, p. 375.

the individuals, Lazarsfeld provides the following set of accounting equations:

$$p_j = \int_{-\infty}^{+\infty} f_j(x) g(x) dx$$

$$p_{jk} = \int_{-\infty}^{+\infty} f_j(x) f_k(x) g(x) dx$$

$$p_{jkl} = \int_{-\infty}^{+\infty} f_j(x) f_k(x) f_l(x) g(x) dx$$

$$\vdots \qquad \qquad \qquad \vdots$$

In the latent distance model, the regression, or trace line, for the items  $j$  is assumed to be a step function of the form shown in Figure 2. It is assumed that the probability of a response is some constant value ( $a_j - b_j$ ) up to a particular point  $x_j$ , and at this point it jumps to another value ( $a_j + b_j$ ) and remains constant thereafter. The equation for the regression can be written

$$\begin{aligned} f_j(x) &= a_j - b_j, & \text{for } x \leq x_j, \\ f_j(x) &= a_j + b_j, & \text{for } x > x_j. \end{aligned}$$

Each item is thus represented by three parameters:  $a_j$ ,  $b_j$ , and a break point,  $x_j$ .

It is a general rule that step-function regression lines provide no metric information about the shape of the underlying distribution of subjects. Any monotonic transformation of the underlying continuum leaves the form of such regression lines unchanged; hence, the final scale is ordinal. Solutions for the latent distance model have been provided by Hays and Borgatta (1954).

It is interesting to note that if the parameters  $a_j$  and  $b_j$  are both assumed to have values of  $\frac{1}{2}$  for all items, the latent distance model reduces to Guttman's scalogram model for dichotomous items.

Lazarsfeld's linear model assumes that the regression lines can be represented by an equation of the form

$$f_j(x) = a_j + b_j x,$$



where  $0 \leq f_j(x) \leq 1$  for all items over the region of the continuum containing the distribution of subjects. Solutions for the item parameters  $a_j$  and  $b_j$  are available (see Torgerson 1958).

**Spearman's two-factor model.** Spearman's two-factor model (1927) differs from Lazarsfeld's linear model only in the form of the manifest data. When it is assumed that the stimulus variables are measured on an interval scale and standardized with a mean of zero and unit variance, then one can write the equation

$$s_{ij} = c_{ij}a_{jg} + e_{ij},$$

where  $s_{ij}$  = score of subject  $i$  on stimulus  $j$ ,  $c_{ij}$  = score of subject  $i$  on the general factor  $g$ ,  $a_{jg}$  = loading of stimulus  $j$  on the general factor, and  $e_{ij}$  represents variance due to unique factors and to error.

The regression line can be written simply as

$$f_j(x) = a_jx,$$

which differs from Lazarsfeld's linear model only by the lack of an additive constant.

**Multiple factor analysis.** The standard procedures of multiple factor analysis require manifest data on interval scales and differ from Spearman's model by the substitution of a multidimensional latent space for the unidimensional general factor. The basic equation for the score of subject  $i$  on stimulus  $j$  becomes

$$s_{ij} = \sum_{m=1}^M c_{im} a_{jm} + e_{ij},$$

where  $m = 1, 2, \dots, M$  is an index for dimensions. The corresponding regression function becomes

$$f_j(x) = \sum_{m=1}^M c_{im} a_{jm}.$$

[See FACTOR ANALYSIS; the biography of SPEARMAN.]

**Normal ogive and logistic models.** Normal ogive and logistic models require dichotomous items scored either as 0 or as 1, and assume a unidimensional latent attribute. They differ from the latent distance and linear models in their assumption of regression lines that seem intuitively to be more reasonable, particularly for measuring aptitudes and abilities. Models using the normal ogive trace lines have been developed by Lord (1953) and Tucker (1952). Birnbaum (1957) and Rasch (1960) have worked with the logistic function.

WARREN S. TORGERSON

[Directly related are the entries LATENT STRUCTURE; MATHEMATICS; PROBABILITY; PSYCHOMETRICS. Other relevant material may be found in FACTOR ANALYSIS;

the articles listed under MEASUREMENT; PSYCHOPHYSICS; QUANTAL RESPONSE; and in the biographies of FECHNER; SPEARMAN; THURSTONE; WEBER, ERNST HEINRICH.]

#### BIBLIOGRAPHY

- ANDERSON, T. W. 1954 On Estimation of Parameters in Latent Structure Analysis. *Psychometrika* 19:1-10.
- ATTNEAVE, FRED 1949 A Method of Graded Dichotomies for the Scaling of Judgments. *Psychological Review* 56:334-340.
- BIRNBAUM, A. (1957) 1958 *Probability and Statistics in Item Analysis and Classification Problems*. Texas: U.S. Air Force School of Aviation Medicine, Randolph Air Force Base.
- BRADLEY, RALPH A. 1954 Incomplete Block Rank Analysis: On the Appropriateness of the Model for a Method of Paired Comparisons. *Biometrics* 10:375-390.
- BRADLEY, RALPH A.; and TERRY, MILTON E. 1952 Rank Analysis of Incomplete Block Designs. I. The Method of Paired Comparisons. *Biometrika* 39:324-345.
- CLARKE, FRANK R. 1957 Constant-ratio Rule for Confusion Matrices in Speech Communication. *Journal of the Acoustical Society of America* 29:715-720.
- CLIFF, NORMAN 1959 Adverbs as Multipliers. *Psychological Review* 66:27-44.
- COOMBS, CLYDE H. 1950 Psychological Scaling Without a Unit of Measurement. *Psychological Review* 57:145-158.
- COOMBS, CLYDE H. 1953 Theory and Methods of Social Measurement. Pages 471-535 in Leon Festinger and Daniel Katz (editors), *Research Methods in the Behavioral Sciences*. New York: Dryden.
- COOMBS, CLYDE H. 1964 *A Theory of Data*. New York: Wiley.
- EDWARDS, ALLEN L.; and THURSTONE, L. L. 1952 An Internal Consistency Check for Scale Values Determined by the Method of Successive Intervals. *Psychometrika* 17:169-180.
- GARNER, WENDELL R.; and HAKE, HAROLD W. 1951 The Amount of Information in Absolute Judgments. *Psychological Review* 58:446-459.
- GIBSON, W. A. 1959 Three Multivariate Models: Factor Analysis, Latent Structure Analysis, and Latent Profile Analysis. *Psychometrika* 24:229-252.
- GREEN, BERT F. JR. 1951 A General Solution for the Latent Class Model of Latent Structure Analysis. *Psychometrika* 16:151-166.
- GULLIKSEN, HAROLD 1954 A Least Squares Solution for Successive Intervals Assuming Unequal Standard Deviations. *Psychometrika* 19:117-139.
- HAYS, DAVID G.; and BORCATT, EDGAR F. 1954 An Empirical Comparison of Restricted and General Latent Distance Analysis. *Psychometrika* 19:271-279.
- LAZARSFELD, PAUL F. 1954 A Conceptual Introduction to Latent Structure Analysis. Pages 349-387 in Paul F. Lazarsfeld (editor), *Mathematical Thinking in the Social Sciences*. Glencoe, Ill.: Free Press.
- LORD, FREDERIC M. 1953 An Application of Confidence Intervals and of Maximum Likelihood to the Estimation of an Examinee's Ability. *Psychometrika* 18:57-76.
- LUCE, R. DUNCAN 1959 *Individual Choice Behavior: A Theoretical Analysis*. New York: Wiley.
- MOSTELLER, FREDERICK 1951 Remarks on the Method of Paired Comparisons. Part 3: A Test of Significance

- for Paired Comparisons When Equal Standard Deviations and Equal Correlations Are Assumed. *Psychometrika* 16:207-218.
- RASCH, GEORG W. 1960 *Probabilistic Models for Some Intelligence and Attainment Tests*. Copenhagen: No publisher listed.
- SAFFIR, MILTON A. 1937 A Comparative Study of Scales Constructed by Three Psychophysical Methods. *Psychometrika* 2:179-198.
- SPEARMAN, CHARLES E. 1927 *The Abilities of Man: Their Nature and Measurement*. London: Macmillan.
- STEVENS, S. S. 1966 A Metric for the Social Consensus. *Science New Series* 151:530-541.
- STEVENS, S. S.; and GALANTER, E. H. 1957 Ratio Scales and Category Scales for a Dozen Perceptual Continua. *Journal of Experimental Psychology* 54:377-411.
- STOUFFER, SAMUEL A. et al. 1950 *Measurement and Prediction. Studies in Social Psychology in World War II, Vol. 4*. Princeton Univ. Press. → See especially the chapters by Louis Guttman and by Paul F. Lazarsfeld.
- SUPPES, PATRICK; and ZINNES, JOSEPH L. 1963 Basic Measurement Theory. Volume 1, pages 1-76 in R. Duncan Luce, Robert R. Bush and Eugene Galanter (editors), *Handbook of Mathematical Psychology*. New York: Wiley.
- THURSTONE, L. L. 1927 A Law of Comparative Judgment. *Psychological Review* 34:273-286.
- THURSTONE, L. L.; and CHAVE, ERNEST I. (1929) 1937 *The Measurement of Attitude: A Psychophysical Method and Some Experiments With a Scale for Measuring Attitude Toward the Church*. Univ. of Chicago Press.
- TORGERSON, WARREN S. 1958 *Theory and Methods of Scaling*. New York: Wiley.
- TORGERSON, WARREN S. 1961 Distances and Ratios in Psychophysical Scaling. *Acta psychologica* 19:201-205.
- TUCKER, L. R. 1952 A Level of Proficiency Scale for a Unidimensional Skill. *American Psychologist* 7:408 only.

## SCHELER, MAX

Max Scheler (1874-1928), German philosopher and sociologist, was born in Munich. His father came from an upper-middle-class Protestant family which can be traced to the sixteenth century; a number of his forebears had been clergymen and jurists in the city of Coburg. His mother was descended from a Jewish Orthodox family that had been settled in Franconia for many centuries. The clash of religious and cultural traditions in his home may account in part for the strains and tensions in his personality and work.

Scheler studied medicine and philosophy at the University of Jena. There, Rudolf Eucken, the Nobel prize-winning idealistic philosopher and defender of Protestant cultural liberalism, was his most prominent teacher. Scheler's early work, beginning with his dissertation at Jena (1899), where

he became a *Privatdozent* in 1901, clearly derived from Eucken. The distinctive aspects of Scheler's thought became apparent only after he moved to the University of Munich in 1907. There he assimilated the phenomenological method: Franz Brentano, who had been Edmund Husserl's teacher, continued to lecture there in Scheler's day, and Scheler was also influenced by Husserl's eminent disciples at the university. Although he was never an "orthodox" phenomenologist, his subsequent philosophizing moved within the orbit of Brentano's and Husserl's thought. [See the biography of HUSSERL.]

In 1910 Scheler gave up his connection with the University of Munich and moved to Berlin to live as an academically unattached writer. His first major works were written in this Berlin period. Among his close intellectual companions in those days were Walther Rathenau and Werner Sombart. The outbreak of World War I marked a turning point in Scheler's career. He threw himself with great ardor into the defense of the German cause. The articles and books in which he defended the "German war" with passionately nationalistic fervor brought him to the attention of a wider public than had noticed his more scholarly productions. In 1917 and 1918 he worked at various diplomatic and propagandistic tasks for the German Foreign Office in Geneva and The Hague. During the war years Scheler, despite his intense nationalistic commitment to the "fatherland," moved closer to a Christian position in the realm of ethics, becoming finally a Roman Catholic convert.

In 1919 Scheler accepted a call to the chair of philosophy and sociology at the University of Cologne. During the next five years he gradually developed his sociology of knowledge. About 1924 he began slowly to turn away from his previous commitment to Catholicism and eventually left the church. He then began to work on a comprehensive *Anthropologie*, in which he attempted to expound a kind of vitalistic pantheism that he had come to adopt. But he was not able to develop this new conception fully. He accepted a call from the University of Frankfurt in the beginning of 1928, but he died on May 19, 1928, at the age of 54, before assuming his new post.

Quite different strands of thought were woven into the texture of Scheler's thought at different times in his intellectual development. He was ever open to new ideas and did not fear to contradict himself. As a consequence, he failed to achieve a full synthesis of his thought but was able to appeal by virtue of his own intellectual restlessness to the restless young intellectuals of the postwar period.



An erratic man, he was always to be a somewhat disturbing figure to the more settled members of the academic community while he gained considerable influence among younger sociologists and philosophers. Although he did not found a "school," he was a major intellectual pathsetter in pre-Nazi Germany as a social critic and moralist and as a sociologist and philosopher. Since World War II the vogue of phenomenology and existentialism in France has led to vigorous interest there in Scheler's work. In the Anglo-Saxon countries, Scheler was until recently little known, except among scholars interested in the sociology of knowledge or among certain theologians and philosophers. The recent translation of certain of his writings has stimulated wider interest among social scientists.

**Work in social psychology.** Scheler's major contributions to the social sciences lie in the domain of social psychology and the sociology of knowledge. His first major work, *Ressentiment* (1912), contains a comprehensive phenomenological description of this sentiment as well as an attempt to locate the *ressentiment*-laden man within the social structure. He developed the notion that certain social roles—the spinster, the mother-in-law, the priest, for example—predispose their occupants to *ressentiment*, an attitude arising from a sense of impotence in the face of the cumulative repression of feelings of hatred, envy, and desire for revenge. In contrast to Nietzsche, who coined the term, Scheler located the breeding ground of *ressentiment* in the modern bourgeois and post-bourgeois world rather than in Christianity.

In *The Nature of Sympathy* (1913), Scheler presented a detailed description of this feeling state. Here he wedded the phenomenological method to the Pascalian endeavor to outline a "logic of the heart." Scheler attempted to uncover eternal uniformities in feelings and emotions and to show that these, far from being the blind results of mechanistically operating associations, are actually means of knowing which reveal, through their intentionality, the situation of man in the universe and the ethical a priori of a distinct realm of eternal values. As in much of his later work, Scheler defended the thesis that values exist independently of the men who make the evaluations and justified his resolute opposition to all pragmatist, naturalist, or positivist theories of value. Polemics against such philosophical currents, more especially against Kantianism and Neo-Kantianism, form a major part of Scheler's work. His *Formalismus in der Ethik* . . . —first published in 1916 but written before the war (*Gesammelte Werke*, vol. 2)—contains the most elaborate development

of his anti-Kantian polemics as well as ethical speculations and sociological considerations that anticipate many of his later sociological ideas. *On the Eternal in Man* (1921) is the fullest exposition of Scheler's religious thought and also contains germs of his later sociological views.

**The sociology of knowledge.** Scheler's contribution to the sociology of knowledge is to be found in his "Probleme einer Soziologie des Wissens" (1924), two years later expanded in *Die Wissensformen und die Gesellschaft* (*Gesammelte Werke*, vol. 8). He attempted to synthesize a Platonic doctrine of eternal immutability of a world of value essences with a comprehensive relativism. He described how different groups of men have striven, each in its socially and historically limited way, to grasp aspects of the eternal sphere of value essences. The infinite variety of subjective a priori, the fact that different groups, or periods, or individual types elaborate their own forms of knowledge, meant to Scheler that men are striving to attain the value essences in different ways at different times rather than that the immutability of these very essences is limited. *Real factors* (such as biological, political, or economic constellations) favor or oppose the actualization of *ideal factors* (for example, moral, religious, or intellectual values), but they can never determine their content. They can only act as "sluice gates of the spirit." Scheler rejected Comte's "law of three stages" and stressed that religious, metaphysical, and scientific knowledge do not succeed each other in regular progression but rather coexist in every age, rooted though they are in the conditions of life of different groups of men and different human types. Yet Scheler found a grain of truth in Comte's dogmatic assertion that there are stages in human history in the sense that in different periods different substructural elements determine the predominant cultural outlook of the age. At the dawn of history, "blood"—that is, racial and vital factors—chiefly controlled cultural life. After the emergence of the state, political factors moved to the foreground. With the bourgeois age, economic factors have achieved unquestioned predominance. While earlier periods accorded pre-eminence to religious and metaphysical types of knowledge, the bourgeois age is primarily the age of the scientist.

Scheler's world view has a deeply aristocratic cast. The idea of hierarchy is central to it, and the equalitarian tendencies of the modern age were therefore deeply abhorrent to him. His doctrine of moral values is hierarchical: pleasure values, those concerning the pleasant or unpleasant, are inferior to vital values, those promoting well-being and

health, while these are again inferior to spiritual values; the highest values, however, are of a religious or sacred character. To such a hierarchy of values corresponds a hierarchy of men representing them. The saint incarnates the sacred values at the very top of the hierarchy; the genius stands for the spiritual values; the hero stands for the vital values; even in the lowest world, that of pleasure, "artists of consumption" are needed to guide our uncertain taste.

To the hierarchy of values and of men embodying them there corresponds a hierarchy of forms of sociation. The lowest form is the simple horde, in which emotional contagion and unconscious imitation constitute the only bond. In the vital community, on the other hand, in the family or tribe as well as in classes or professions, conscious solidarity binds the participants. As against the horde, the social here prevails over the individual; each member can easily be substituted for any other. In the third type of sociation, Tönnies' *Gesellschaft*, superficial contractual bonds between individuals replace supra-individual goals that are consciously supported. Finally, at the apex of the hierarchy, the *Gesamtperson*, the "complex collective personality," represents a type of sociation in which a quintessential solidarity binds the members in a community of love. This last form, of which the church and the nation are central examples, is eminently superior to all others; within it alone can the higher values be realized.

To posit equality between persons and so to level essential differences of value was, to Scheler, the chief aberration of the modern age. His full-scale attack against the post-Christian world, his rejection of liberalism, altruism, humanitarianism, and democracy can best be understood as resulting from his conviction that the modern world with its leveling tendencies has undermined the objective hierarchy of men and values. His cultural criticism hence proceeds from a profoundly elitist point of view. The modern social scientist will of necessity have to sift Scheler's very real contributions to sociology and social psychology from the antidemocratic ideology in which they are all too often embedded.

LEWIS A. COSER

[For the historical context of Scheler's work, see KNOWLEDGE, SOCIOLOGY OF; and the biographies of COMTE; HUSSERL; KANT; TÖNNIES. For discussion of the subsequent influence of Scheler's ideas, see EXPRESSIVE BEHAVIOR; PERCEPTION, article on PERSON PERCEPTION; PERSONALITY, POLITICAL; PHENOMENOLOGY; SYMPATHY AND EMPATHY; and the biography of MANNHEIM.]

#### WORKS BY SCHELER

- 1899 *Beiträge zur Feststellung der Beziehungen zwischen den logischen und ethischen Prinzipien*. Jena: Voelkelius. → Inaugural dissertation, University of Jena.
- (1912) 1961 *Ressentiment*. Edited with an introduction by Lewis A. Coser. New York: Free Press. → First published as *Über Ressentiment und moralisches Werturteil*. The second German edition appears on pages 33-147 of Volume 3 of Scheler's *Gesammelte Werke* under the title "Das Ressentiment im Aufbau der Moralen."
- (1913) 1954 *The Nature of Sympathy*. Translated by Peter Heath, with an introduction by Werner Stark. London: Routledge. → First published as *Zur Phänomenologie und Theorie der Sympathiegefühle und von Liebe und Hass*. The second German edition was published as *Wesen und Formen der Sympathie*.
- (1921) 1961 *On the Eternal in Man*. Translated by Bernard Noble. New York: Harper. → First published as *Vom Ewigen im Menschen*.
- 1924 *Probleme einer Soziologie des Wissens*. Pages 5-146 in Max Scheler (editor), *Versuche zu einer Soziologie des Wissens*. Munich and Leipzig: Duncker & Humblot.
- (1927) 1961 *Man's Place in Nature*. Translated with an introduction by Hans Meyerhoff. New York: Farrar, Boston: Beacon. → First published in German in Volume 8 of *Der Leuchter*.
- (1929) 1958 *Philosophical Perspectives*. Translated by Oscar A. Haac. Boston: Beacon Press. → First published as *Philosophische Weltanschauung*.
- Gesammelte Werke*. Vols. 1—. Bern (Switzerland): Francke, 1954-. → Thirteen volumes are projected, of which the following have appeared: Volume 2: *Der Formalismus in der Ethik und die materiale Wertethik: Neuer Versuch der Grundlegung eines ethischen Personalismus* (1913-1916) 1954. Fourth edition, revised. Volume 3: *Vom Umsturz der Werte: Abhandlungen und Aufsätze* (1911-1914) 1955. Fourth edition, revised. Volume 5: *Vom Ewigen im Menschen* (1921) 1954. Fourth edition, revised. Volume 6: *Schriften zur Soziologie und Weltanschauungslehre* (1923-1924) 1963. Second edition, revised. Volume 8: *Die Wissensformen und die Gesellschaft* (1926) 1960. Second edition, revised. Volume 10: *Schriften aus dem Nachlass* (1933) 1957. Second edition, revised and enlarged.

#### SUPPLEMENTARY BIBLIOGRAPHY

- BECKER, HOWARD; and DAHLKE, HELMUT OTTO 1942 Max Scheler's Sociology of Knowledge. *Philosophy and Phenomenological Research* 2:310-322.
- COSER, LEWIS A.; and HOLDREIM, WILLIAM W. 1961 *Max Scheler: An Introduction*. New York: Free Press.
- DUPUY, MAURICE 1959 *La philosophie de Max Scheler: Son évolution et son unité*. 2 vols. Paris: Presses Universitaires de France.
- GURVITCH, GEORGES D. (1930) 1949 L'intuitionnisme émotionnel de Max Scheler. Pages 67-152 in Georges Gurvitch, *Les tendances actuelles de la philosophie allemande*: E. Husserl, M. Scheler, E. Lask, N. Hartmann, M. Heidegger. Paris: Vrin.
- KANTHACK, KATHARINA 1948 *Max Scheler: Zur Krisis der Ehrfurcht*. Berlin and Hannover: Minerva.
- MCGILL, V. J. 1942 Scheler's Theory of Sympathy and Love. *Philosophy and Phenomenological Research* 2: 273-291. → The entire issue is devoted to Scheler.



- MANNHEIM, KARL (1925) 1952 *Sociology of Knowledge From the Standpoint of Modern Phenomenology* (Max Scheler). Pages 154-179 in Karl Mannheim, *Essays on the Sociology of Knowledge*. New York: Oxford Univ. Press. → Contains a discussion of the differences between Mannheim's and Scheler's approaches to the sociology of knowledge. First published in German.
- STARK, WERNER 1958 *The Sociology of Knowledge: An Essay in Aid of a Deeper Understanding of the History of Ideas*. London: Routledge; Glencoe, Ill.: Free Press.
- STAUDE, JOHN R. 1967 *Max Scheler: An Intellectual Portrait*. New York: Free Press.

## SCHIZOPHRENIA

Schizophrenia constitutes the core problem of insanity. The conditions subsumed under the heading are among the most devastating to which man is heir, blighting him in the process of development or in the prime years of life and commonly leading to his progressive enclosure in an autistic world beset by delusion, hallucination, and fear of involvement with life and people. Because of the diverse manifestations, courses, and outcomes, schizophrenic reactions elude precise definition and concise description. They may be defined as gross failures to achieve or maintain integrated personality functioning in which a person, unable to cope with problems of living and irreconcilable conflicts, withdraws and seeks solutions by breaking through the confines imposed by the culture's ways of thinking and reasoning and by regressing to a period of childhood when reality gave way to fantasy and when the self and others were not yet clearly distinguished. The condition tends to be self-perpetuating because the patient abandons testing his ideas in terms of how they help master his environment and promote collaborative interaction with others. However, restitutive efforts are usually made because a person cannot achieve an equilibrium as an isolate and because fears of uncontrollable impulses interfere with autistic gratification.

**Incidence and prevalence.** Schizophrenia is among the most common conditions leading to profound incapacitation. Of the daily hospital census in the United States of about 1.4 million patients, approximately one-fourth are hospitalized because of schizophrenia. Although the large majority are discharged within the first year after admission, the average length of stay is 13 years because a sizable number remain in hospitals from adolescence or early adulthood until death from old age or some disease. Incidence and prevalence can be estimated only very roughly because of the

number of schizophrenic patients who are never hospitalized or even diagnosed and because of the vague boundaries of the condition. It has been estimated that between 14 and 20 of every 1,000 children born in the United States will be hospitalized at some time in their lives with an illness diagnosed as schizophrenia. A minimum annual prevalence rate of 290 per 100,000 has been suggested for Western societies (Lemkau & Crocetti 1958). Surveys in New Haven (Hollingshead & Redlich 1958) and in England (Brooke 1959) have shown that both incidence and prevalence are much higher among the lower socioeconomic segments of the population than among the upper-middle and upper classes. [See MENTAL DISORDERS, article on EPIDEMIOLOGY.]

## Concepts of schizophrenia

The profound divergencies of opinion concerning the nature, etiology, and treatment of schizophrenic conditions reflect differing concepts of the nature of human adaptation and integration more than they reflect divergent findings. Traditionally, psychiatrists have considered that the profound incapacitations of schizophrenic patients and, particularly, the disordered thinking that forms a critical aspect of the condition must be due to some unknown biological impairment—a structural defect of the brain or some hereditary metabolic or toxic factor influencing cerebral functioning. A large majority continues to believe that the high familial incidence indicates a hereditary predisposition that is affected by environmental factors, perhaps through aggravating a metabolic dysfunction. However, others consider schizophrenic reactions as an extreme type of aberrant personality functioning resulting from faulty and distorting child rearing in deviant family environments.

The way in which schizophrenia is conceptualized is not simply a matter of academic interest but of vital moment, as it influences the direction of scientific investigations and therapeutic efforts and, as will become apparent, has even profoundly influenced the clinical picture and course of the condition. The conviction that schizophrenia is a disease due to a biological impairment leads to concentration of research in neuropathological, physiological, and biochemical fields and has created therapeutic pessimism. This attitude has often contributed to the neglect of patients and, at times, virtually condemned them to crowded and impersonal institutions that aggravated their autistic withdrawal. Such orientation left little room for the participation of social scientists, other than in studies of the economic and social implications

of a devastating disease. In contrast, the conceptualization of schizophrenia as a type of aberrant personality development has made it possible to consider schizophrenic reactions in the study of how the developmental process can go awry and thus perhaps to clarify critical issues concerning human integration and emotional homeostasis. Investigations of such matters as the role of language in human adaptation, of the family as a critical social institution that mediates between biological and cultural directives and between the individual and society, of the influence of social class on the incidence of mental illness and how it relates to child-rearing practices, of cross-cultural influences on personality development, and of the social and political structure of mental hospitals and the society's system of caring for the ill, and numerous other areas of study become a vital part of the effort to understand, control, and treat schizophrenic reactions. Further, as schizophrenia constitutes one of the dominant unsolved health problems, the training of all physicians as well as of psychiatrists in particular will be affected profoundly by the conceptualization of the problems and the projection of an approach to their solution.

**Historical considerations.** The concept of schizophrenia as a clinical entity derives from Emil Kraepelin's recognition in 1896 of the relationship between three forms of insanity—catatonia, described by K. Kahlbaum in 1863; hebephrenia, described by E. Hecker in 1871; and the deteriorating paranoid psychoses—and his uniting them under the designation "dementia praecox," which B. Morel had used for a type of deteriorating adolescent psychosis. Kraepelin brought a degree of order to psychiatric taxonomy by differentiating dementia praecox from the manic-depressive psychoses. He emphasized the progressive course of dementia starting in youth and envisioned a distinct disease entity for which a specific structural pathology or causative agent would be found. The term "schizophrenia" was introduced by Eugen Bleuler in 1911, because not all sufferers from dementia praecox were youthful and not all became demented; it also served to emphasize the splitting of the psychic functions—the intellectual, affective, and conative—in these patients. Bleuler considered the possibility of a psychological causation but favored a hereditary etiology. He considered that the unknown etiologic agent produced primary symptoms of disturbances in associations, abnormalities of affect, emotional and intellectual ambivalence, and turning from reality to fantasy, as well as autism, whereas such symptoms as delusions, hallucinations, motor disorders, etc., were secondary and

amenable to psychological treatment. He also stressed the intact capacities for perception, memory, and intellectual functioning that differentiate schizophrenia from the organic and toxic psychoses. In essence, he outlined the critical paradox of schizophrenic conditions: thought disorders form a critical aspect and diagnostic requisite, but the anatomical and physiological capacities for perceiving, thinking, and remembering remain intact. Together with Carl Gustav Jung, Bleuler brought psychoanalytic insights into the study of the psychology of schizophrenia. [See the *biographies* of BLEULER; JUNG; KRAEPELIN.]

**Adolf Meyer's view.** For the vast majority of psychiatrists throughout the world, the schizophrenic syndrome has remained essentially unchanged since the publication of Bleuler's monograph (1911), although many have adhered more closely to Kraepelin's concepts of a clear-cut disease entity, as the collected papers of the 1957 World Congress devoted to schizophrenia demonstrate (World Congress . . . 1959). However, Adolf Meyer, the dominant American contemporary of Bleuler, did not consider schizophrenia as an illness but as a reaction type that could be brought on by faulty "habit patterns," including habits of thinking and relating (Meyer 1896-1937). For many years Meyer was almost the only important influence advocating longitudinal dynamic studies of life histories of schizophrenic patients, and he countered the prevailing therapeutic pessimism with his melioristic approach which advocated a type of psychotherapeutic management in a socializing milieu. Meyer is largely responsible for the different perspectives concerning schizophrenia held by many American psychiatrists in contrast to those of their Continental colleagues. [See the *biography* of MEYER.]

**Psychoanalysis.** Even though Freud had excluded the treatment of schizophrenia from the province of psychoanalysis because he tacitly accepted an organic or toxic etiology and considered that the regression to narcissism prevented formation of a transference relationship, psychoanalysis still made many major contributions to understanding schizophrenic patients and their symptoms. Such concepts as narcissistic fixation and regression, the psychosexual stages of development, infantile sexuality, the withdrawal of cathexes (libidinal investment) from objects, primary process thinking, the interpretation of dreams, the exposition of the defense mechanisms—particularly those of projection and introjection—and of the importance of homosexuality to paranoid states, etc., opened the way for a dynamic under-



standing of schizophrenic patients and for comprehension of productions that previously had seemed to be the incomprehensible utterances of a diseased mind. Harry Stack Sullivan was among the first analysts who focused attention on the treatment of schizophrenic patients and whose psychoanalytic orientation was greatly modified by work with them. In the process he found need and use for contributions from various behavioral sciences. Sullivan (1924-1933) and his colleague Fromm-Reichmann (1948) were major forces in demonstrating by determined efforts that psychoanalytic therapy could be adapted to the treatment of schizophrenic patients. [See the biography of SULLIVAN.]

### Clinical descriptions

The nature of the disorder varies widely. Most often schizophrenic reactions become manifest in the adolescent or young adult when the impact of sexual drives and the attainment of independence, an identity, and a way of life present critical problems. The young person may gradually and insidiously become increasingly withdrawn, preoccupied, and resistant to direction from others. He may avoid facing problems and relating to others by sleeping through much of the day. Sudden outbursts of irritability or hostility toward family members ward off intrusions. Rumination over personal and philosophical problems increases, and intellectualizations displace commitment and show of emotions. Differentiation from adolescent "identity crises" or other types of adolescent turmoil is difficult, and the future of the youth may hang in the balance. Gradually withdrawal becomes more extreme, indecision paralyzing, and ambivalence of feelings pronounced. Appearance and bodily hygiene may be neglected. Communication becomes vague, missing the mark and leaving the listener puzzled. Feelings of unreality and depersonalization are common. Self-direction is rescinded or paralyzed, allowing drive, impulse, and fantasy to gain increased motivating force. Control of drives and hostilities becomes a major preoccupation, while motivation is sought through interpreting coincidental events and reading meanings into people's expressions and gestures. Then delusional solutions and hallucinated directives are found. Unacceptable feelings and impulses are projected onto others, who become malevolent figures. Hypochondriacal complaints are common and may reflect fears of transformation into the opposite sex. Increasingly the patient lives in an autistic world, gaining a sense of worth or even greatness through fantasy; he is also a prey to delusions of persecution and to terror lest homicidal, suicidal,

or incestuous impulses overwhelm him. In some patients the first serious indication of the psychosis comes in an outbreak of panic over loss of control or over projected dangers. Frequently, when the patient can remain isolated or when he is neglected in an institution, habits and bodily care deteriorate, communications become disorganized and bizarre, and the patient comes to live in a world of delusion and hallucination. He exists in a state of almost living death, a nonparticipant in society and its conventions, virtually beyond the reach of therapeutic intervention. However, such outcomes are not inherent to schizophrenic reactions and need not occur, and with improved treatment they are becoming less common even in large institutions. The course and outcome vary widely and are influenced markedly by the treatment provided. Under average current conditions, approximately two-thirds of hospitalized patients will be discharged within 12 months: about half of these will remain reasonably well, while the remainder will have further serious difficulties and tend to have a downhill course. The outlook is poor for those who cannot be discharged within the first year. However, as less flagrant forms of schizophrenia are now diagnosed more often than formerly and as better treatment opportunities are available to some, the diagnosis need not imply so poor a prognosis. Schizophrenic reactions are most likely to occur in persons who have socialized poorly and who tend to be shy, introspective, somewhat eccentric, and, perhaps, compliantly overconscientious. Still, a fair proportion have not been notably schizoid, and some have been outgoing but impetuous individuals. Lifelong difficulties in socializing and an early progressive withdrawal decrease chances of recovery.

**Types of schizophrenia.** Conventionally, four types of schizophrenia are described. Actually, most patients can be fitted into these types only arbitrarily, showing admixtures of two or more types, with shifts in the dominant symptomatology over time. However, brief outlines of these subgroupings will serve to describe some of the different clinical pictures.

**Catatonic schizophrenia.** In catatonic schizophrenia extremes of violent motor excitement or rigid immobility dominate the picture. Onset is often abrupt. The patient may maintain difficult postures for hours, days, or months and requires complete care. Alternations between extreme excitement and rigidity may occur. Frequently these patients are in the midst of some mystical experience, believing themselves in heaven or hell; they are often immobile and refuse to speak because

they believe any movement or word can produce a universal catastrophe. Although they may appear out of contact, they are aware of and sensitive to their surroundings. The prognosis is more favorable than for other types but unpredictable in the individual case. Chronic catatonic states, once fairly common, are becoming infrequent in modern hospitals.

*Simple schizophrenia.* Simple schizophrenia usually refers to a gradual withdrawal of interest and a progressive decline of responsible behavior with absence of commitment to a definite way of life. The potentialities shown in youth dissipate, and the patient idles about the home or becomes a vagrant, etc. Theoretically, Bleuler's primary symptoms are present with minimal secondary symptoms; however, many patients given this diagnosis are delusional but noncommunicative about their delusions and hallucinations.

*Hebephrenic schizophrenia.* Hebephrenic schizophrenia usually starts in adolescence and progresses rapidly: the patient soon displays silly, impulsive, and disorganized behavior. Delusions and hallucinations are poorly organized and shifting. Speech is often fragmented and almost incomprehensible, marked by frequent intrusions of primary process material.

*Paranoid schizophrenia.* In paranoid schizophrenia delusions of a persecutory type dominate thought and behavior; ideas of reference are common and blend with auditory hallucinations. Megalomaniac delusions relate to ideas of being the focus of widespread plots. The later in life the onset, the more likely it is that the delusions will be systematized and the reaction type approach the condition described under paranoid states. [See PARANOID REACTIONS.]

*Other subcategories.* The current official nomenclature includes a schizoaffective type to cover the numerous patients who display an admixture of schizophrenic and manic-depressive symptoms: if these are considered separate entities, such patients would be suffering from both conditions. However, if schizophrenic reactions are regarded as a type of aberrant development leading to failures of personality integration, other such combined diagnoses will be used. Sociopathic youths can also be more or less schizophrenic; hypochondriacal neuroses may shift to psychoses with somatic or hypochondriacal delusions; obsessive patients can decompensate with obsessive thoughts becoming increasingly bizarre and delusional. Further, although the typical schizophrenic reactions are readily diagnosed, there are no clear boundaries that delimit the use of the diagnosis.

Those who are reluctant to relinquish the concept of schizophrenia as a progressive deteriorating disease seek to distinguish between process and nonprocess schizophrenia or between schizophrenia and schizophreniform conditions (Langfeldt 1933): the patients who recover or who are amenable to psychotherapy, etc., are not considered as true schizophrenics and supposedly can be distinguished by their premorbid behavior, heredity, and symptomatology—a view not widely shared in the United States.

The diagnosis of pseudoneurotic schizophrenia (Hoch & Polatin 1949) is used by some psychiatrists to indicate patients with panneurotic symptomatology who are not amenable to psychotherapy because they are really schizophrenic, but others consider these as "borderline" patients whose neurotic symptoms form a defense against personality disorganization. Borderline schizophrenia refers to patients who maintain a tenuous integration with meager defenses against incursions of primitive impulses and strange primary process material into consciousness.

Childhood schizophrenia refers to cases occurring before adolescence. It is not clear if cases occurring early in childhood are related to the adult condition. Early infantile autism is a puzzling illness that occurs in the first few years of life; the child begins to display peculiar repetitive behavior and regresses, paying little attention to other persons. Some of these children clearly suffer from brain anomalies or brain damage; others have been raised by unempathic, intellectualizing parents. These conditions cannot be discussed adequately in this section. [See MENTAL DISORDERS, article on CHILDHOOD MENTAL DISORDERS.]

The confusions concerning nomenclature have been presented because the various terms are in common usage but also because they reflect the difficulties and ambiguities in the field. Terms may be used in the literature with a certainty that can be misleading, leaving the inexperienced person—and even the experienced one—feeling ignorant rather than perplexed when he cannot make a clear-cut diagnosis.

### Concepts of etiology

*Impairment of brain functioning.* The conviction that such profound personality disorganizations, marked by disordered thinking and often progressing to an almost vegetative state of existence, must reflect impairment of brain functioning has led to intensive investigations of post-mortem and biopsy sections of the brain and of every endocrine, toxic, and biochemical factor that might



conceivably affect cerebral functioning. The erroneous assumption that insulin coma therapy cured schizophrenia led to an upsurge of such studies in the 1930s and 1940s, and currently the meliorating effects of the phenothiazine drugs have provided renewed impetus. Investigators have again and again reported some biochemical factor they consider specific to schizophrenia, and each of these findings has been invalidated sooner or later. A number of these factors have turned out to be related to such things as dietary inadequacies or are concerned with physiological disturbances secondary to chronic emotional disturbances or to prolonged inactivity or some other influence of chronic hospitalization.

*Neuroanatomical investigations.* None of the numerous reports of an abnormality of the brain has been validated. The similarities of the personalities and developmental histories of patients with psychomotor epilepsy and schizophrenia form an interesting lead that is under study.

*Endocrine investigations.* Although the thyroid, pituitary, gonadal, pineal, and adrenal cortical and medullary secretions have all been implicated, no evidence has held up. The intriguing hypothesis that epinephrine undergoes an anomalous breakdown into psychotoxic substances is being pursued (Hoffer 1964), but recent evidence does not bear out the initial enthusiasm.

*Biochemical investigations.* Recently emphasis has been directed to the study of brain amines and indole excretion, and to psychogenic agents such as LSD-25 and the supposedly psychogenic extracts of schizophrenic serum. Various theories relating several such factors have been proposed. Intensive work is advancing knowledge of brain chemistry and fosters hope for future discoveries pertinent to schizophrenia, but recent premature enthusiasms indicate a need for more carefully controlled and validated work before publication. [See MENTAL DISORDERS, *article on BIOLOGICAL ASPECTS.*]

*Genetic hypothesis.* The high familial incidence of schizophrenia had led to virtual acceptance of the belief that some genetically transmitted factor plays an important role in the etiology of schizophrenia [see MENTAL DISORDERS, *article on GENETIC ASPECTS*]. It has been primarily the evidence from twin studies that seemed to demand acceptance of a strong hereditary influence. Until very recently all twin studies reported a much higher concordance rate in identical twins than in same-sexed fraternal twins or siblings—as high as 86 per cent in identical as against 17 per cent in same-sexed fraternal twins. Lately, several investigators

noted puzzling inconsistencies in the data. The only two studies of all twins born within a given span of time rather than of samples of hospitalized patients fail to validate the earlier findings. A survey of all male twins born in Finland between 1930 and 1935 found that none of the 16 schizophrenic patients among the identical twins had a schizophrenic co-twin (Tienari 1963). Studies in Norway of an even larger sample, although still incomplete, indicate that concordance rates for identical twins will not be much higher than for same-sexed siblings (Kringlen 1964). It appears as if the striking results of prior twin studies were due to sampling errors and that this basis for the genetic hypothesis of schizophrenia has been badly shaken.

*Family environment.* A high familial incidence of personality traits or illnesses has often led to erroneous overemphasis of genetic factors. Attention has been directed increasingly to the role of the family environment in producing schizophrenic offspring. Psychoanalytic theory, in considering schizophrenia related to fixations at the oral stage of development, drew attention to failures in the earliest mother-child transactions. As early as 1924, Sullivan emphasized the noxious influence of certain mothers on the development of their schizophrenic sons, both directly and because of their attitudes toward their husbands (1924-1933). Considerable etiologic significance has been given to the "schizophrenogenic mother" who is either aloof, unempathic, and unable to cathect in the child properly or, more typically, cannot establish boundaries between herself and the child, engulfing and controlling a child she needs to complete her life. Studies since 1949 (Lidz et al. 1965) have revealed deficiencies and abnormalities in the total intrafamilial environment. The fathers are frequently just as severely disturbed as the mothers. The family is split by enduring conflict between the parents in which each demolishes the worth of the other to the children, or the family transactions are distorted because one parent acquiesces to the spouse's strange ways of rearing children and patterning the family life. The generation roles within the family are confused in a variety of ways such as parental rivalry with a child or parental dependency upon an immature child, including heterosexual or homosexual incestuous proclivities. Parents fail to adhere to their respective sex-linked roles, either because of homosexual tendencies or through reversing maternal and paternal roles, or because a mother cannot fill an affectional-expressive role or a father an instrumental role. The parent of the same sex as the child who becomes schizophrenic

does not form an adequate model for identification, a situation that is often aggravated because this parent's worth is undercut by the spouse.

Communication within the family is always disturbed and often clearly irrational or paralogical. The intrafamilial culture may deviate markedly from that of the society into which the child must eventually emerge. These families teach or indirectly inculcate irrationality, providing a poor foundation in reality testing and for understanding verbal and nonverbal communications outside of the family. The schizophrenic patient's siblings are almost always seriously affected: more are psychotic than are reasonably well-integrated, and with rare exception their occasional "normality" is achieved at the price of serious constriction of the personality.

The finding that schizophrenic patients always grow up in seriously disturbed families forms the most consistent lead concerning the etiology of schizophrenia. Essentially similar phenomena have been reported by research groups in many different countries. In conceptualizing how such disturbed family transactions can lead to schizophrenia in an offspring, some investigators focus primarily upon how the disturbed patterns of communication foster the schizophrenic thought disorder. The "double bind" hypothesis (Bateson et al. 1956) notes how the patient is habitually caught in a bind because a parent covertly sends conflicting messages that perplex and paralyze, causing him to be rebuffed whichever way he responds; or he is caught between the opposing needs and demands of his two parents, and satisfying either parent provokes rejection by the other. Various other communication problems have been noted: the ways in which the child is taught to deny or ignore what should be obvious; the distortions of meanings to support a parent's tenuous emotional equilibrium; the fostering of distrust in the utility of verbal communications; the teaching of eccentric or delusional beliefs; the blurring of the system of meanings and constructs by inconsistent reinforcement by parents, etc. (Lidz et al. 1965). The amorphous or fragmented nature of the thinking and communicating of one or both parents has been carefully documented (Wynne & Singer 1963). Studies of the family communication patterns clarify why the patient suffers from a thought disorder and how such resultant impairments in thinking and communicating create profound disturbances in ego functioning.

The family disturbances, however, clearly affect the child's development deleteriously in other significant ways. A broader approach to the problem

takes cognizance of how failures of parental nurturance, particularly the mother's difficulties in relating to the child, interfere with the development of adequate autonomy; how the disturbances in the family structure distort the structuring of the child's personality by improper channeling of drives, through confusing child and parent roles, by creating confusions in sexual identity, by impeding proper resolution of the oedipal situation, etc.; as well as of how parents' failure to transmit adequately basic adaptive or instrumental techniques of the culture, including its system of meanings, impairs ego functioning and socializing capacities (Lidz et al. 1965).

The family studies have thus far been largely exploratory, paving the way for the formulation of hypotheses that are now being studied with more rigorous methodology. Although not definitive, they have served to move research concerning the etiology of schizophrenia out of the frustrating whirlpool in which it had been caught. Many puzzling aspects of schizophrenia now seem far more comprehensible. The findings interdigitate with those of other approaches, such as studies of the relationship between social class and incidence of schizophrenia, twin studies, and psychoanalytic concepts of developmental dynamics. The knowledge gained from these studies has also had a marked impact upon therapy, particularly through drawing attention to the need to consider the family as a unit rather than simply to focus upon the individual patient.

## Therapy

Despite the absence of any specific method of treatment, marked changes have taken place in the therapy of schizophrenic patients since 1940. Perhaps the most important change has been the gradual but progressive abandonment of the defeatist attitude concerning schizophrenia that had pervaded most of psychiatry since the mid-nineteenth century. Although debates have waged concerning the value of insulin coma, electroshock, frontal lobotomy, milieu therapy, and various forms of individual and group psychotherapy, psychiatrists and other personnel finding something they could do or try to do became interested in the patients, and the patients responded. It became apparent to increasing numbers of psychiatrists that schizophrenic patients need not follow a downhill course, and that many of the extreme manifestations were products of neglect and virtual abandonment in impersonal institutions. The enormity of the problem of caring for and attempting to treat the masses of institutionalized patients



has been overwhelming, and adequate treatment for any other than selected patients has had to await reorganizations of hospital systems and the training of the necessary personnel.

Therapeutic efforts must be suited to the patient, the facilities available, and the therapist's abilities. Whereas the prognosis for schizophrenic reactions collectively is generally discouraging, the outlook for the individual patient, particularly if treatment begins soon after symptoms appear and intensive care can be provided, may be regarded hopefully.

**"Organic" treatments.** Although insulin coma therapy, introduced in 1933, did much to revive interest in the treatment of schizophrenic patients, it has now been virtually discarded, its apparent efficacy having resulted from the increased attention to patients it required. Electric convulsive treatments are of occasional value as a means of quieting extremely excited patients. Various tranquilizing drugs, particularly the phenothiazines, have largely replaced other "organic" treatments. The phenothiazines serve to lessen anxiety and agitation and probably diminish distraction by extraneous stimuli and primary process intrusions. Trifluoperazine may be particularly useful in diminishing hallucinations. An array of psychotropic drugs is now available, and opinions differ as to the merits of each. Properly controlled experiments on the relative worth of these medications are extremely difficult to conduct. It seems reasonably certain that many patients can return to or remain in the community largely because of these agents. A major influence of tranquilizing drugs has been indirect; they have helped quiet the wards in mental institutions and provided a means of controlling seriously disturbed patients. This has helped produce great changes in mental hospitals through unlocking doors, permitting patients greater freedom, and allowing the staff time and opportunity to improve the therapeutic milieu. [See MENTAL DISORDERS, TREATMENT OF, article on SOMATIC TREATMENT.]

**Milieu therapy.** Establishing a suitable therapeutic hospital milieu has altered the prognosis for schizophrenic patients as much as the tranquilizing drugs, but in all except certain select institutions, reorganization had to await the quieting effects of the shock therapies and then the tranquilizers. Milieu therapy seeks to counter schizophrenic patients' tendencies to withdraw; it also fosters socialization, promotes responsibility, and provides retraining in interpersonal relationships. The therapy includes a gamut of measures ranging from discarding restraints; minimizing isolation; giving attentive care and interest; fostering socializing, edu-

cational, and occupational activities; and providing group therapy to holding patient-staff meetings and setting up patient government to provide channels of communication and to encourage responsibility for the self and others. Opening the doors of hospitals has a salutary influence and can markedly change the attitudes of both patients and staff, but some patients who are afraid of their impulsivity, aggressions, or suicidal tendencies feel more secure and can socialize more readily when protected within a limited area until they are ready to assume responsibility for their actions.

As schizophrenic patients tend to regress, withdraw, and become passively dependent in institutions, there has been a movement to return such patients to the community as soon as feasible or to attempt to keep them out of hospitals altogether. On the other hand, efforts are often made to promote early hospitalization, for the chances of recovery are greater if treatment starts before symptoms become well established and if the patient can be removed from a pathogenic environment. The paradox revolves around the nature of the treatment a hospital can offer; optimally a hospital should be able to counter regressive tendencies. As yet, relatively few institutions can provide intensive care and treatment that seeks to promote essential personality change rather than mere remission from a psychotic state.

**Psychotherapy.** Establishing a relationship to another person often—perhaps always—provides the impetus and forms the bridge for the schizophrenic patient's return from psychotic withdrawal. The crux of the initial phases of psychotherapeutic work with these patients is the therapist's use of himself to gain the patient's trust and willingness to risk relating again despite the shattering disillusionments of the past. A persistent warm and honest interest by an untrained person may suffice and succeed where a highly trained but intellectualized and uncommitted therapist fails. Schizophrenic patients are usually highly sensitive to pretense, to being used by another person, and to another person's withdrawal in the face of their intense needs or their hostilities. Usually, however, considerable skill and understanding of schizophrenic patients are required. Trust follows upon understanding, understanding requires communication, and the therapist faces the task of establishing communication across barriers imposed by the patient's idiosyncratic language usage, his personalized metaphor, his efforts to conceal and yet convey, the delusions and hallucinatory interruptions, and the projections of the patient's thoughts and feelings onto the therapist, etc. The therapist

must be prepared and able to counter or weather the patient's flights into withdrawal, particularly when the patient feels on the verge of trusting and becoming involved—setbacks that test the therapist's commitment. Bringing a patient out of his psychotic regression and withdrawal can be a very real and rewarding achievement, but something that often occurs without specific psychotherapy; the process is more successful when the therapist is a focal point in a proper therapeutic milieu that counters the patient's regression in many ways than when the emphasis is solely upon the individual psychotherapy.

Psychotherapy of schizophrenia usually implies more than bringing the patient out of the psychotic episode or into socially acceptable behavior. It often involves the effort to promote profound personality changes through psychoanalytically oriented interpersonal transactions. Such work with schizophrenic patients differs markedly from the analysis of psychoneurotic patients. It seeks to strengthen ego functioning through fostering a more cohesive identity, firmer boundaries between the self and others, and control of primitive impulses. Treatment almost always includes extensive re-evaluation of parental figures and efforts to modify the sway of sabotaging parental introjects. The intense nature of the transference relationship once it is established, together with the fragility of the relationship, places great demands upon the therapist. It is a task for highly skilled therapists with specific experience and training in the field and will not be discussed here. The effectiveness of intensive psychotherapy with schizophrenic patients cannot be evaluated in statistical terms: the belief in its value derives from the excellent results obtained with a number of individuals and from the meaningfulness of the material gained in the process. Much depends on the specific therapist as well as on the specific patient. The aims of such work are directed in part toward the future—to increasing knowledge and improving techniques. [See CLINICAL PSYCHOLOGY; MENTAL DISORDERS, TREATMENT OF, especially the article on PSYCHOLOGICAL TREATMENT; PSYCHIATRY.]

**Family therapy.** Recent advances in understanding of the pathology in the schizophrenic's family of origin have brought the family into the therapeutic efforts in a variety of ways. Social workers or psychiatrists have worked with parents individually or in groups to allay their anxiety and guilt; to counter the common tendency of parents to interfere with or disrupt the patient's therapy; to modify the parents' attitudes toward each other as well as toward the patient; and to foster more

direct and forthright communication. With some families a major effort is made to prevent exclusion of a hospitalized patient and withdrawal of interest in him and in others to help prevent disintegration of the family after the patient has been hospitalized. Conjoint family therapy in which parents and patient or the entire family are seen in group sessions has been found advantageous. The therapist has an opportunity to view directly the nature of the disturbed relationships and communications; the patient may gain a new perception of his parents and an appreciation of their limitations and idiosyncrasies; and other family members may recognize their own roles in provoking or continuing the patient's illness, gain insights concerning the family transactions, and begin to relate differently. Although some psychiatrists consider conjoint family therapy to be the optimal treatment procedure, others use it as an important adjunct to alternative therapeutic measures.

### Perspective

The disagreements between virtually all schools of psychiatry concerning the nature and etiology of schizophrenia have interfered with the presentation of a coherent and cohesive approach to the problem. Although schizophrenia is traditionally considered a disease of unknown etiology, a great deal is now known about the developmental problems and the dynamic psychopathology of these patients and about the family environments that tend to produce them. Man is unique in the degree that his adaptation rests on a lengthy period of dependency upon parental figures during which he must assimilate the instrumental techniques of his culture to implement his inborn adaptive capacities. There are countless chances for misdirection, confusion, and conflict to arise in the process. The very mechanism that permits his inordinate adaptability contains a major vulnerability. He depends upon language and a coherent way of thinking to guide him into the future and on an ability to communicate to enable collaborative interaction. Yet linguistic meanings and a system of logic are not inborn but acquired as a means of problem solving by communication with others and by sorting out life experiences. How correctly meanings are learned and how firmly they become ingrained depend largely on the parental tutors—the consistency of their communications as an aid to problem solving and the emotional relationships between parents and child. There are countless ways in which the enculturation process can go wrong. When a person is caught in insoluble conflicts, when a path into the future is barred, when



even regression serves little because the persons upon whom one could depend are distrusted or feared, there is still a way. One can alter the perception of his own needs and motives and those of others. One can abandon causal logic, change his internal representation of events, retreat to a period of childhood when reality gave way to fantasy, and cut off movement toward a realistic future; that is, one can become schizophrenic. This path is so clearly open to man, particularly to those who have had confused or confusing guides to follow into adulthood and have received poor foundations in meaning systems and reality testing, or who have actually been trained to irrational ways in childhood, that conditions such as schizophrenia must be expected as anomalies of the human developmental process. The acceptance of such a theory that schizophrenic reactions result from anomalous socialization processes and that the family forms the primary enculturating agency provides numerous guides for research and therapy, and it directs the psychiatrist toward increased collaboration with behavioral scientists.

THEODORE LIDZ

[Directly related are the entries MENTAL DISORDERS; PSYCHOSIS. Other relevant material may be found in MENTAL DISORDERS, TREATMENT OF; PSYCHIATRY; PSYCHOANALYSIS; and in the biographies of KRAEPELIN; MEYER; SULLIVAN.]

#### BIBLIOGRAPHY

- ALANEN, Y. O. 1958 The Mothers of Schizophrenic Patients: A Study of the Personality and the Mother-Child Relationship of 100 Mothers and the Significance of These Factors in the Pathogenesis of Schizophrenia in Comparison With Heredity. *Acta psychiatrica scandinavica* 33 (Supplement 124).
- ARIETI, SILVANO 1955 *Interpretation of Schizophrenia*. New York: Brunner. → A readable and cogent presentation of a modern dynamic approach.
- BATESON, GREGORY et al. 1956 Toward a Theory of Schizophrenia. *Behavioral Science* 1:251-264.
- BELLAK, LEOPOLD (editor) 1958 *Schizophrenia: A Review of the Syndrome*. New York: Logos. → Reviews work in the field between 1946 and 1956; contains 4,000 references.
- BLEULER, EUGEN (1911) 1950 *Dementia Praecox: Or, the Group of Schizophrenias*. New York: International Universities Press. → Still an outstanding work on the topic. Includes an extensive bibliography to 1950. First published as "Dementia Praecox oder die Gruppe der Schizophrenien" in Aschaffenburg's *Handbuch der Psychiatrie*.
- BRODY, EUGENE B.; and REDLICH, F. C. (editors) 1952 *Psychotherapy With Schizophrenics*. New York: International Universities Press.
- BROOKE, EILEEN M. 1959 National Statistics in the Epidemiology of Mental Illness. *Journal of Mental Science* 105:893-908. → Now called the *British Journal of Psychiatry*.
- CAMERON, NORMAN A. 1938 *Reasoning, Regression and Communication in Schizophrenics*. Psychological Monographs. Vol. 50, no. 1. Columbus, Ohio: Ohio State University. → A classic study of schizophrenic thinking.
- DELAY, JEAN; DENIKER, P.; and GREEN, A. 1957-1962 *Le milieu familial des schizophrènes*. 3 parts. *Encéphale* 46:189-232; 49:1-21; 51:5-73. → Part 1: Position du problème, 1957. Part 2: Méthodes d'approche, 1960. Part 3: Résultats et hypothèses, 1962.
- FREEMAN, THOMAS; CAMERON, JOHN L.; and MCGHIE, ANDREW 1958 *Chronic Schizophrenia*. New York: International Universities Press.
- FROMM-REICHMANN, FRIEDA 1948 Notes on the Development of Treatment of Schizophrenics by Psychoanalytic Psychotherapy. *Psychiatry* 11:263-273.
- HOCH, PAUL; and POLATIN, PHILLIP 1949 Pseudoneurotic Forms of Schizophrenia. *Psychiatric Quarterly* 23:248-276.
- HOFFER, ABRAM 1964 The Adrenochrome Theory of Schizophrenia: A Review. *Diseases of the Nervous System* 25:173-178.
- HOLLINGSHEAD, AUGUST B.; and REDLICH, F. C. 1958 *Social Class and Mental Illness: A Community Study*. New York: Wiley.
- JACKSON, DON D. (editor) 1960 *The Etiology of Schizophrenia*. New York: Basic Books. → An excellent collection of papers on various contemporary theories.
- KASANIN, J. S. (editor) 1944 *Language and Thought in Schizophrenia*. Berkeley and Los Angeles: Univ. of California Press. → Brief articles by various authorities that can serve as an introduction to the topic.
- KRINGLEN, EINAR 1964 Schizophrenia in Male Monozygotic Twins. *Acta psychiatrica scandinavica* 40 (Supplement 178):1-76.
- LANGFELDT, G. 1933 Some Points Regarding the Symptomatology and Diagnosis of Schizophrenia. *Acta psychiatrica et neurologica scandinavica Supplement* 80:7-26. → Now called *Acta psychiatrica scandinavica*.
- LEMEAU, PAUL V.; and CROCETTI, GUIDO M. 1958 Vital Statistics of Schizophrenia. Pages 64-81 in Leopold Bellak (editor), *Schizophrenia: A Review of the Syndrome*. New York: Logos.
- LIDZ, THEODORE; FLECK, STEPHEN; and CORNELISON, ALICE R. 1965 *Schizophrenia and the Family*. New York: International Universities Press. → Collected papers of an intensive study of the intrafamilial environment of schizophrenic patients.
- MEYER, ADOLF (1896-1937) 1951 *Collected Papers. Volume 2: Psychiatry*. Baltimore: Johns Hopkins Press. → For the origins of the genetic-dynamic approach, see the chapters "Dementia Praecox" and "Paranoia" on pages 413-560.
- REES, W. LINFORD 1957 Physical Characteristics of the Schizophrenic Patient. Pages 1-14 in Derek Richter (editor), *Schizophrenia: Somatic Aspects*. London: Pergamon.
- Die Schizophrenie. 1932 Volume 9, special part 5 in Karl Wilmanns (editor), *Handbuch der Geisteskrankheiten*. Berlin: Springer. → An authoritative source for the German approach in the Kraepelinian tradition. Contains an outstanding bibliography.
- SINGER, MARGARET T.; and WYNNE, LYMAN C. 1965a Thought Disorder and Family Relations of Schizophrenics: III Methodology Using Projective Techniques. *Archives of General Psychiatry* 12:187-200.
- SINGER, MARGARET T.; and WYNNE, LYMAN C. 1965b Thought Disorder and Family Relations of Schizo-

- phrenics: IV. Results and Implications. *Archives of General Psychiatry* 12:201-212.
- SULLIVAN, HARRY STACK (1924-1933) 1962 *Schizophrenia as a Human Process*. New York: Norton.
- Symposium international sur la psychothérapie de la schizophrénie. 1957 *Acta psychotherapeutica, psychosomatica et orthopaedagogica* 5:99-360.
- SYMPOSIUM ON SCHIZOPHRENIA, SAN FRANCISCO, 1958 1959 *Schizophrenia: An Integrated Approach*. Edited by Alfred Auerbach. New York: Ronald Press.
- TENARI, P. 1963 Psychiatric Illnesses in Identical Twins. *Acta psychiatrica scandinavica* 39 (Supplement 171). → The whole supplement is devoted to Tenari's study.
- WORLD CONGRESS OF PSYCHIATRY, SECOND, ZURICH, 1957 1959 *Report*, 4 vols. Zurich: Füssli.
- WYNNE, LYMAN C.; and SINGER, MARGARET T. 1963 Thought Disorder and Family Relations of Schizophrenics. 2 parts. *Archives of General Psychiatry* 9: 191-206. → Part 1: A Research Strategy. Part 2: A Classification of Forms of Thinking.

### SCHLESINGER, KARL

Karl Schlesinger (1889-1938) was born in Budapest and died in Vienna, committing suicide on the day Hitler took possession of that city. (He had moved there after Béla Kun's 1919 communist revolution in Hungary.) A wealthy financier and a member of many industrial and financial boards, Schlesinger was one of the very few important contemporary theoretical economists who was not part of the academic world; he never held an academic post but was an active and highly respected member of the Vienna Economic Society. Although his contributions to economics were greatly appreciated in Vienna, his important early work of 1914, *Theorie der Geld- und Kreditwirtschaft* (see 1914a), was not readily available at the time he moved there; its direct influence upon Viennese monetary theory was less than might have been expected.

In his 1914 book Schlesinger made extensive use of some simple mathematics, uncommon at that time in German economic writing, particularly in the field of monetary theory. His interest in mathematics led him to take instruction from A. Wald in the 1930s; he eventually proposed to Wald a problem relating to Walras' production equations. This problem concerned Schlesinger's modification of the Walras-Cassel system of production equations by introducing the notion that *all* factors of production—not only the scarce ones—have to be considered and that whether a factor is scarce or free depends on the demand functions. This led to Wald's famous work on the determinacy of the Walrasian system. [See ECONOMIC EQUILIBRIUM.] (Schlesinger acknowledged that Zeuthen had earlier

suggested a similar modification without, however, following up the matter.) These problems were discussed in Karl Menger's mathematical colloquium under Menger's continuing and stimulating guidance. Schlesinger also contributed to Menger's work on the theory of returns.

Schlesinger's *Theorie der Geld- und Kreditwirtschaft* made him the only immediate follower of Walras, other than Wicksell, to advance Walras' theory of money. Schlesinger stressed that in addition to the indirect utility of money derived from its ability ultimately to provide desired consumers' goods, cash balances also have a direct utility. He introduced a probabilistic approach to monetary theory by distinguishing between fixed and uncertain future due dates for payments; the former offer no choice, while the latter permit the individual to maximize utility by choosing appropriately between holding cash and purchasing goods. He further showed that if the number of transactions increases, it is possible to economize in the relative size of cash balances, although this will not be the case if only the magnitude of transactions increases. Schlesinger derived an excess demand equation for money that is virtually identical with the one commonly ascribed to Keynes. [See MONEY.] He also was probably the first to develop the notion of the equilibrium rate of interest [see INTEREST].

Schlesinger was one of the first economists to expound the purchasing power parity theory. In his pamphlet on the variations in the value of money (1916), he stated the theorem that the parity of a country's currency is determined by the quotient of the domestic and foreign general price levels. The equilibrium rate of exchange can only be changed if foreign borrowings are used to reduce the indebtedness of the bank of issue.

In 1931 Schlesinger published a lengthy review of F. Somary's *Bankpolitik* (see 1931a), in which he presented a detailed mathematical analysis of the theory of credit creation by commercial banks, incorporating this theory into the general process of circulation. Assuming there are  $n$  equal banks, that deposits require some cash reserve, and that each individual pays partly in cash and partly by check, Schlesinger analyzed the effects of introducing new metal or new bank notes into the system. He traced the path that credit expansion follows until equilibrium is reached and derived formulas for debtors, creditors, and the system as a whole. He also considered the flow of checks back to their starting point. Several of his restrictive assumptions can be relaxed, but the whole approach is an early model for careful, rigorous analysis of monetary operations. [See BANKING.]



During the international financial crisis of the 1930s Schlesinger proposed that the Austrian government control capital movements, but he was strictly opposed to exchange control as then practiced.

Schlesinger's most significant contributions were to monetary theory, but this work was widely neglected, which caused Schumpeter to comment that "in our field first-class performance is neither a necessary nor a sufficient condition for success" (1954, p. 1082, note 5). More recently, however, he has been given more credit, especially by V. Wagner and D. Patinkin. In addition to his analytical contributions to the numerous discussions about capital theory, business cycles, and general equilibrium theory conducted among economists in Vienna—frequently at odd hours in coffee houses—Schlesinger had important economic insights in which his great experience as a financier and administrator provided a valuable counterweight to his theoretical analysis. Thus, his considerable influence on Viennese economics was far greater than his few publications might indicate.

OSKAR MORGENSTERN

[See also the biographies of WALD and WALRAS.]

#### WORKS BY SCHLESINGER

- 1914a *Theorie der Geld- und Kreditwirtschaft*. Munich: Duncker & Humblot.  
 (1914b) 1959 Basic Principles of the Money Economy. *International Economic Papers* 9:20–38. → This is a partial translation of *Theorie der Geld- und Kreditwirtschaft*.  
 1916 *Die Veränderungen des Geldwertes im Kriege*. Vienna: Manz.  
 1931a [A Book Review of] *Bankpolitik*, by Felix Somary. *Archiv für Sozialwissenschaft und Sozialpolitik* 66: 1–35.  
 1931b Das "Rätsel" der französischen Goldpolitik. *Zeitschrift für Nationalökonomie* 2:387–407.  
 1935 Über die Produktionsgleichungen der ökonomischen Wertlehre. *Ergebnisse eines mathematischen Kolloquiums* 6:10–11.  
 1936 Ein Beitrag zu Mengers "Bemerkungen zu den Ertragsgesetzen." *Zeitschrift für Nationalökonomie* 7: 398–400.

#### SUPPLEMENTARY BIBLIOGRAPHY

- ELLIS, HOWARD S. 1934 *German Monetary Theory, 1905–1933*. Harvard Economic Studies, Vol. 44. Cambridge, Mass.: Harvard Univ. Press.  
 MARGET, ARTHUR W. 1931 Léon Walras and the "Cash-balance Approach" to the Problem of the Value of Money. *Journal of Political Economy* 39:569–600.  
 PATINKIN, DON (1956) 1965 *Money, Interest, and Prices: An Integration of Monetary and Value Theory*. 2d ed. New York: Harper.  
 SCHUMPETER, JOSEPH A. (1954) 1960 *History of Economic Analysis*. Edited by E. B. Schumpeter. New York: Oxford Univ. Press.

WAGNER, VALENTIN F. 1937 *Geschichte der Kredittheorien: Eine dogmenkritische Darstellung*. Vienna: Springer.

#### SCHLICK, MORITZ

Friedrich Albert Moritz Schlick (1882–1936), philosopher and epistemologist was born in Berlin. He studied physics at Heidelberg and Lausanne and took his doctorate at Berlin under Max Planck with a dissertation on the physics of light. He then qualified himself for an academic post in philosophy and taught first at Rostock, then Kiel, and finally at Vienna from 1922 to 1936, where he occupied the chair in the philosophy of the inductive sciences, held earlier by Ernst Mach and Ludwig Boltzmann. Schlick became the leading figure of the so-called Vienna circle—a group of thinkers concerned with developing a scientific philosophy in opposition to the prevailing metaphysical orientation of German philosophical thought.

After a brief but highly productive period this school was disrupted and scattered, first by Schlick's death on June 22, 1936 (he was murdered by a mentally disturbed former student), and then by the Nazi invasion of Austria in 1938. However, its influence continued and was strongly felt for another two decades, especially in English-speaking countries (where it became known as logical positivism), as a consequence of the emigration of some of its leading members and sympathizers, among them Rudolf Carnap, Hans Reichenbach, and the somewhat less closely connected Karl Popper.

Although trained as a physicist and strongly motivated by the desire to bring philosophical clarity to the knowledge of nature and to render philosophy more scientific in method and results, Schlick was sensitive to the whole range of problems of knowledge, human existence, and values. He had a simple, clear style free of all technical jargon, rising on occasion to something like the level of poetry.

His first book, *Lebensweisheit* (1908; "Wisdom of Life"), as the name indicates, was devoted to nonscientific issues, as was his last, *Problems of Ethics* (1930a), which, while largely an attempt to give an account of ethics from the causal-explanatory point of view, contains many profound observations on life and values. Aside from these two books and a small separately published essay on the meaning of life, *Vom Sinn des Lebens* (1927), Schlick's contributions to philosophy fall in the fields of the philosophy of science and epistemology.

He did not, however, concern himself with the philosophy of the social sciences; at least he published nothing dealing directly with its special problems (however, see *Natur und Kultur*). The most important of his ideas pertaining to this area are those that are contained in his ethical writings mentioned above. A few words will be said about them before we turn to his principal contributions dealing with knowledge as such.

### Moral philosophy

Schlick distinguished sharply between cognition and experience. Reality is given to us in experience, but that experience is not itself knowledge. This distinction, between *Erlebnis* and *Erkenntnis* ("experience" and "cognition"), which plays so fundamental a role in his theory of knowledge, is no less important in his ethics and *Lebensphilosophie*. Thus, the moral philosopher seeks only to understand moral phenomena, moral conduct—not to guide or create them. The moralist, wise man, or prophet, on the other hand, is interested in the quality of life itself, seeks to enlighten men with respect to the good life, and thus has a practical, not theoretical, goal.

As a moralist, Schlick taught that the good life is the life full, to the highest degree, of activity valued for its own sake. Activities valued in and for themselves are properly called play, and Schlick agreed with Schiller that man is truly man only when he plays. And, as youth is characteristically the time of play, the concept of "youth" is the clue to the meaning of life. Whenever man acts out of the joy in acting, apart from considerations of ends to be achieved, then he is, regardless of his age, youthful in the philosophical sense, and then his life is meaningful. "If we need a rule for life let it be: 'Preserve the spirit of youth!', for it is the meaning of life" (1927, p. 354).

As a moral philosopher, Schlick viewed his task as that of explaining moral conduct, i.e., how it comes about that man does what is required of him by society, which is the moral lawgiver making demands upon him to act, even when such conformance might seem contrary to his own desires. Normative ethics—defining "the good," "what ought to be"—he considered as at most a preliminary to ethics. It is not the task or privilege of the moral philosopher to decide what society ought to mean by these terms but only to ascertain what it does mean. But this does not imply that the philosopher docilely accepts as good whatever society holds to be such; his task is to understand and explain norms, not to accept or reject them. Schlick's view was that society deems and calls good that which

it supposes is beneficial to it—i.e., makes for its happiness—and it demands that its members act accordingly. The ultimate ethical norms, if such there be, are facts in human experience and, precisely because they are ultimate, they neither need nor admit of justification—a senseless requirement but one all too frequently made in the history of ethical thought. Schlick wrote that "even if ethics were a normative science it would not cease because of this to be a science of facts . . ." ([1930a] 1939, p. 21).

According to Schlick, in a situation of choice a man selects that goal the idea of which at the time of choice is relatively more pleasure-tinged or less tinged with unpleasant feeling than the alternatives. Thus men can, and often do, knowingly choose goals that are in their realization unpleasant. In the long run, however, the motive pleasure is in turn affected by the realization pleasure in such a way that goals regularly found to be unpleasant upon achievement, whether because of their own nature or because of social sanctions, cease to be attractive. The combined effect of man's natural sympathy and of social sanctions is to make man identify his own happiness with that of his fellow men. Thus arises that close connection between virtue (moral conduct, i.e., altruism) and happiness, which explains why men act morally.

Of possible interest to students of society is Schlick's analysis of the concept of "responsibility," which he thought should replace the traditional concept of so-called freedom of the will, a concept that in his view is irrelevant to ethics and based on a series of confusions, namely, between causality (determinism) and compulsion, indeterminism (chance) and freedom. That a man's actions are determined means only that they are predictable, not that they are coerced. "Compulsion" occurs where a man is prevented from realizing his natural desires. How could the rule according to which these natural desires arise itself be considered as 'compulsion'?" (*ibid.*, p. 148). Morality and society are both concerned only with who is responsible for the acts they wish to discourage or encourage, and this influence would be impossible if acts were free in the sense of undetermined, i.e., uncaused.

### Philosophy of science and epistemology

Schlick's work in the theory of knowledge and the philosophy of science dominated his scholarly career. It falls into two periods: one culminating in the publication of his chief work, *Allgemeine Erkenntnislehre* (1918), and another beginning with his move to Vienna in 1922 and ending with his death in 1936.



**Experience and knowledge.** The most important single idea of *Allgemeine Erkenntnislehre* is that the experience of something (*Erlebnis*) is not the same as the knowledge of it (*Erkenntnis*). To know something is basically to recognize it, i.e., to find in it something already familiar. Schlick frequently made use of the pair of German terms *kennen* and *erkennen* to make this distinction; the former refers to being acquainted with something, the latter to recognizing it as an instance of something else. This recognition ranges from the simplest cases, such as finding a shade to be a certain color, to recognizing in the phenomena of light something governed (described) by the same laws as those governing electrical phenomena. The whole goal of knowledge as *theory* is to achieve conceptual unification, to discover identity amid diversity. This is a joy in itself, but in addition it gives man the power to master endlessly diverse phenomena with a minimum of basic concepts.

Knowledge is correct description, a true statement or judgment. Contrary to the teaching of Kant, our knowledge is not confined to the realm of phenomena, what is sensibly given, but reaches over into the realm of "things-in-themselves," if by this term we refer to what is not phenomenon, not given in experience. For we can make true judgments about what we do not and cannot experience, as well as about what we can. Knowledge is symbolic, not intuitive.

Such things-in-themselves are defined not in terms of experienced "content" but implicitly, in terms of their relational properties. To know something about such things is not to intuit their qualities but to be acquainted with the laws that describe their behavior. To be real is not to be simply an object of sensible intuition but rather to occupy a position in time. The moment it is shown that for any object the rules of some particular scientific discipline force us to grant it a definite time and place, its real existence is assured. Thus, the existence and nature of atoms, of fields of force, or of the experiences of other persons are vouched for by the laws that describe them correctly, enabling us to predict and control, not by their entering qualitatively into our experience. Physical space and time are not identical with their sensible (subjective) correlates (visual space is not Euclidean) but are known through the laws that describe them.

Considerations of a similar nature were used by Schlick in his polemic against the phenomenalism and "positivism" of such thinkers as Hume, Avenarius, and Mach, whose doctrines he called the "immanence idea" and subjected to sustained and effective criticism.

**Analysis and verification.** Under the influence of Wittgenstein, with whom he was in close personal contact during his Vienna period, certain aspects of Schlick's thought underwent a gradual but appreciable shift of emphasis. Wittgenstein's conception of philosophy as the *activity* of clarifying meaning furnished Schlick with a clearer line of demarcation between philosophy and science, or positive knowledge: the former clarifies concepts and analyzes the meaning of statements, determining their experiential truth conditions; the latter seeks to verify them, determine their truth values. Philosophical activity constitutes the necessary precondition of science but does not itself put forth doctrines, does not take the form of a body of knowledge. It consists rather in those sometimes verbal acts by means of which meanings are "shown." The ultimate meanings of words and statements must be found in what is given in experience, in content.

The great error of metaphysics has been the effort to express the inexpressible, the content of experience. Here again appears the main theme that knowledge is not mere awareness of the given but concerns its form or structure, which alone is capable of being expressed and communicated. Knowledge describes the *relations* which hold between the elements of content, their structure, and specifically, the laws connecting them.

The meaning of a proposition is given by describing the experiential conditions under which it will be true and those under which it will be false, i.e., the conditions in the given that verify and falsify it. This meaning principle, the verifiability principle, plays the dominant role in the philosophical writings of Schlick's final period. Since all meaningful statements must be verifiable and falsifiable in the given, there are no limits to knowledge, for what transcends possible ascertainment does so by transcending significance itself.

It does not follow from the verifiability that only the given exists. This doctrine is itself senseless, as is the opposite view that the given is unreal, mere appearance. "To be real always means to stand in a definite relationship to the given" ([1932] 1959, p. 99). It is in the given that we verify or falsify (strictly, confirm or disconfirm) statements about existence and reality by inferring, from the hypothesis that something exists, that certain experiences will be forthcoming consequent on certain tests or operations and then testing for the truth of these predictions. The laws on the basis of which we make these predictions are the true objects of science. "The subject matter of physics is *not* sensations but laws. The formulation, used by some positivists, that bodies are only

'complexes of sensations' is therefore to be rejected. What is correct is only that propositions concerning bodies are transformable into equivalent propositions concerning the occurrence of sensations in accordance with laws" (*ibid.*, p. 107).

Schlick, in keeping with his nontechnical manner of writing, did not always give very precise formulation to the verifiability principle. In general he distinguished between analytic (logical, necessary) and synthetic (empirical, contingent) statements. Only the latter, and not all of them, are propositions in the fullest sense of being verifiable and falsifiable. Analytic statements are limiting cases, necessary truths, not expressing possible states of the world but definitional, "grammatical" decisions or rules. The logical is the purely formal, the *a priori*; that there are no synthetic *a priori* truths is a central theme of his polemic against Kant and a view common to the school as a whole.

Statements of laws—general hypotheses—not being strictly verifiable, are not genuine, full-fledged propositions but are treated as rules for making predictions, thus as heuristic principles. Since they are not even theoretically verifiable, the question constituting the problem of induction cannot arise, for laws are not propositions (true or false) whose truth values might be inferred from others. Induction is but a methodical kind of guessing about useful rules.

The question of whether determinism is true, whether *all* events are related to others by regularities expressible in laws, is to be distinguished from the question of the nature of causality. The statement that determinism is true is not verifiable or falsifiable and hence is not a genuine proposition. As a guiding principle, however, determinism might be expressed by the precept "Seek causal laws!" A causal law is a formula enabling us to make successful predictions, given certain data. That it holds necessarily (universally), i.e., that determinism is true within a given domain, cannot be *finally* established or refuted, but we can find good empirical grounds for accepting or rejecting the view in particular cases.

Despite his stress on verifiability and falsifiability, Schlick in fact considered all propositions formulated in daily life and science to be hypotheses, at best confirmable or disconfirmable but not strictly verifiable—at least from the point of view of ascertainment, if not of logical theory. The only contingent statements whose truth can be, in the strict sense, established are those that, on the moment of their utterance, report or announce the given (a here-now-thus) in experience. These confirmations (*Konstatierungen*, as he called them) consti-

tute the ultimate connections between language and the world, "the unshakeable point of contact between knowledge and reality" ([1934] 1959, p. 226). All testing of other propositions is in terms of the data constituting the referents of these *basic* propositions, which immediately cease to have meaning when they lose contact with these referents. The propositions actually made use of in science and daily life are "a means of finding one's way among the facts; of arriving at the joy of confirmation . . ." (*ibid.*, p. 226). The pragmatic nature of Schlick's concepts of knowledge and propositions is thus evident.

**Limits of knowledge.** Since knowledge is of laws describing events and since content is incommunicable and cannot be expressed in propositions, we can know the nature of things only in terms of the laws that apply to them, laws that enable us to make successful predictions in the manner described above. The age-old problem of the relation of mind and body finds its solution here. Mind or, better, the mental is what obeys the qualitative laws of psychology; body, matter, or the physical is what obeys the quantitative laws of physics, which, in contrast to the (nonbehavioristic) laws of psychology, are intersensual and intersubjective. The problem of the connection of mind and body is thus the problem of the connection of the *languages* of psychology and physics.

The as yet unestablished but not implausible hypothesis of *physicalism* (materialism freed of its metaphysical connotations) is that the qualitative aspects of experience (content) are correlated with certain physical, quantitative, measurable properties in such a way that from determination of the latter (brain states, nerve impulses, etc.) universally correct inferences could be made to the former; and in general that a complete description of the world can be made on the basis of physical concepts. If this should be correct, then the mental would be explained by the physical, would in a certain sense *be* the physical. That this is so we do not know, but the problem is insoluble and senseless if it requires explaining how extended substance, i.e., traditional "body," can interact with unextended substance—"mind." The mental is in any case often extended, for example, a pain can be spread over an area or occupy a place; hence the traditional problem disappears, to be replaced by one concerning the empirically determinable relation between certain kinds of propositions, not kinds of substances. And if solved, it will be solved by scientific experimentation and discovery, not by "philosophical" or metaphysical considerations.

The reworking of his central ideas in the manner just illustrated—on the basis of his conception



of knowledge as symbolic, essentially concerned with communication and expression, and of expression as dealing with form, not content—was the task that confronted Schlick when he was cut down so tragically in the prime of life.

DAVID RYNIN

[See also CAUSATION; ETHICS, article on ETHICAL SYSTEMS AND SOCIAL STRUCTURES; POSITIVISM; SCIENCE, article on THE PHILOSOPHY OF SCIENCE.]

#### WORKS BY SCHLICK

- 1908 *Lebensweisheit*. Munich: Beck.  
 (1917) 1920 *Space and Time in Contemporary Physics*. Oxford: Clarendon. → First published in German.  
 (1918) 1925 *Allgemeine Erkenntnislehre*. 2d ed. Berlin: Springer.  
 1927 *Vom Sinn des Lebens*. Sonderdrucke des Symposium, Heft 6. Berlin: Weltkreis-Verlag.  
 (1930a) 1939 *Problems of Ethics*. Englewood Cliffs, N.J.: Prentice-Hall. → First published in German. A paperback edition was published in 1962 by Dover.  
 (1930b) 1959 *The Turning Point in Philosophy*. Pages 53–59 in A. J. Ayer (editor), *Logical Positivism*. Glencoe, Ill.: Free Press. → First published as "Die Wende der Philosophie" in Volume 1 of *Erkenntnis*.  
 (1931) 1961–1962 *Causality in Contemporary Physics*. *British Journal for the Philosophy of Science* 12, no. 47:177–178; no. 48:281–298. → First published in German.  
 (1932) 1959 *Positivism and Realism*. Pages 82–107 in A. J. Ayer (editor), *Logical Positivism*. Glencoe, Ill.: Free Press. → First published in German in Volume 3 of *Erkenntnis*.  
 (1934) 1959 *The Foundation of Knowledge*. Pages 209–227 in A. J. Ayer (editor), *Logical Positivism*. Glencoe, Ill.: Free Press. → First published in German in Volume 4 of *Erkenntnis*.  
*Aphorismen*. Edited by Blanche Hardy Schlick. Vienna: Privately published, 1962. → Published posthumously.  
*Gesammelte Aufsätze: 1926–1936*. Vienna: Gerold, 1938. → Chapters 6, 7, 8, 14, and 15 were written in English. An abridged edition was published in 1948 as *Gesetz, Kausalität und Wahrscheinlichkeit*.  
*Natur und Kultur*. Edited by Josef Rauscher. Vienna and Stuttgart: Humboldt, 1952. → Published posthumously.  
*Philosophy of Nature*. New York: Philosophical Library, 1949. → First published posthumously in German in 1938.

#### SUPPLEMENTARY BIBLIOGRAPHY

- FEIGL, HERBERT 1937/1938 *Moritz Schlick. Erkenntnis* 7:393–419.  
 KRAFT, VICTOR (1950) 1953 *The Vienna Circle: The Origin of Neo-positivism; a Chapter in the History of Recent Philosophy*. New York: Philosophical Library. → First published in German.

#### SCHMIDT, WILHELM

Wilhelm Schmidt was born in 1868 in Hörde, Westphalia, and died in 1954 in Fribourg, Switzerland. He joined the Divine Word Missionaries (*Societas Verbi Divini*, or S.V.D.) in 1883 and

was ordained a priest in 1892. From 1893 to 1895 he studied at the University of Berlin, majoring in Oriental languages; he was appointed professor in the St. Gabriel Mission Seminary at Mödling, near Vienna, in 1895 and remained there until 1938. He also taught at the University of Vienna, beginning in 1921, and after he left Austria, he taught at the University of Fribourg from 1939 to 1951. He founded the periodical *Anthropos* in 1906, was director of the Museo Missionario-Etnológico Lateranense in Rome from 1927 to 1939, and was director of the Anthropos Institut at St. Augustin near Bonn from 1932 to 1950.

Early in his career, Schmidt was very interested in linguistic studies, but later he turned increasingly to studies in ethnology and the science of religion. The impetus to his work in these latter fields came from his philosophical, theological, and linguistic training and from the observations made by S.V.D. missionaries, especially in New Guinea and Togo. He was influenced as well by the folklorist Andrew Lang, by the Indologist Leopold von Schröder, and by the ethnologists Friedrich Ratzel, Leo Frobenius, Ernst Grosse, C. N. Starcke, Alfred Vierkandt, Edward Westermarck, Franz Boas, and, above all, by Fritz Graebner and his theories of *Kulturkreise*.

In his studies of the evolution and forms of the family, Schmidt followed Grosse in correlating family types with economic activities. He refined Graebner's theory of *Kulturkreise*, which differentiated between matrilineal and patrilineal societies, seeking to explain their origins on the basis of subsistence patterns; thus, food planters would be matrilineal in descent and residency rules, while nomads would be patrilineal.

He agreed with others, such as Havelock Ellis and Westermarck, in offering psychological explanations for exogamy. Also, his arrangements of traits in *Kulturkreise* are frequently in accord with functionalist ideas (although he did not want to be associated with "functionalism"—see 1944 for his violent criticism of functionalism). Thus totemism is identified with clan exogamy, in contrast with the consanguineal basis of local exogamy; education is linked to initiation ceremonies, which are usually obligatory for all the young men of the tribe, rather than to secret societies, which often constitute an elite and have special requirements for admission.

Schmidt was concerned with the variability of individuals and their relationship to society. He asserted that the individual, even among primitives, exerts an influence on the institutions of his community.

In his 12-volume work, *Der Ursprung der Gottes-*

*idee* (1912–1955), he made use of extensive material supporting the theory that the Supreme Being occupied a position so superior (compared with spirits and other beings of religious significance) in the religion of primitive hunters and gatherers that this religion may properly be considered monotheistic. Subsequently, such monotheism persisted especially in nomadic pastoral civilizations, while in the *Kulturkreise* of the matrilineal agricultural peoples and the more advanced totemic hunting cultures, the figure of the Supreme Being was stifled or supplanted by other religious elements, e.g., fertility cults, ancestor worship, magic. Totemism itself was regarded by Schmidt as essentially social; he particularly stressed its significance for clan exogamy. Although totemism does imply a certain reverence for the totem animal, Schmidt did not consider it a religious phenomenon—an actual animal cult.

From the time that Schmidt became familiar with Graebner's concept of *Kulturkreise* (Graebner believed that he established such *Kulturkreise* for Oceania, including Australia), this concept exerted an influence on his linguistic work. He tried to show that when Australian languages are classified, different groups coincide with totemistic and matrilineal cultures respectively. Furthermore, he attributed certain linguistic phenomena to economic and social change; thus, he felt that the change to uxorial residence and the ownership of land by women might result in the learning of a foreign tongue by the men, and thus contribute to the disintegration of the language. Indeed, Schmidt made an even more comprehensive attempt to relate all aspects of linguistic evolution to general cultural evolution, to explain the former largely in terms of the latter, and to set up parallels between linguistic areas and *Kulturkreise*.

Schmidt's influence was far-reaching, especially in the German- and English-speaking regions of the world, where his journal *Anthropos* was a repository of field data collected by missionaries around the world. Schmidt was also concerned that his findings have a practical application to colonial policy and missionary practice. Many of his publications, especially those dealing with family and social ethics, were addressed to a public wider than the academic one.

JOSEPH HENNINGER

[For the historical context of Schmidt's work, see ANTHROPOLOGY; and the biographies of BOAS; ELLIS; FROBENIUS; GRAEBNER; RATZEL; WESTERMARCK. For discussion of the subsequent development of Schmidt's ideas, see HISTORY, article on CULTURE HISTORY; LINGUISTICS, article on HISTORICAL LINGUISTICS; and the biography of KOPPERS.]

# WORKS BY SCHMIDT

- 1906 Die moderne Ethnologie; L'ethnologie moderne. *Anthropos* 1:134–163, 318–387, 592–643, 950–997.
- 1908–1910 L'origine de l'idée de Dieu. *Anthropos* 3:125–162, 336–368, 559–611, 801–836, 1081–1120; 4:207–250, 505–524, 1075–1091; 5:231–246. → A revised German version appeared in book form in 1912 as Volume 1 of *Der Ursprung der Gottesidee*.
- 1910 Die Stellung der Pygmäenvölker in der Entwicklungsgeschichte des Menschen. Studien und Forschungen zur Menschen- und Völkerkunde, Vols. 6–7. Stuttgart (Germany): Strecker & Schröder.
- 1912–1955 Der Ursprung der Gottesidee: Eine historisch-kritische und positive Studie. 12 vols. Münster in Westfalen (Germany): Aschendorff. → A second edition of Volume 1 appeared in 1926.
- 1924 SCHMIDT, WILHELM; and KOPPERS, WILHELM Völker und Kulturen. Regensburg (Germany): Habel.
- (1925) 1964 Familie. Pages 13–42 in Wilhelm Schmidt, Wege der Kulturen: Gesammelte Aufsätze. *Anthropos* Institut, Studia, Vol. 20. St. Augustin bei Bonn (Germany): Verlag des Anthropos-Instituts. → First published in the *Handwörterbuch der Staatswissenschaften*.
- 1926 Die Sprachfamilien und Sprachenkreise der Erde. Heidelberg (Germany): Winter.
- (1927) 1946–1949 Rassen und Völker in Vorgeschichte und Geschichte des Abendlandes. 3d ed. 3 vols. Lucerne (Switzerland): Stocker. → First published as *Rasse und Volk*.
- 1930 Handbuch der vergleichenden Religionsgeschichte zum Gebrauch für Vorlesungen an Universitäten, Seminaren usw. und zum Selbststudium: Ursprung und Werden der Religion. Theorien und Tatsachen. Münster in Westfalen (Germany): Aschendorff.
- 1935 The Position of Women With Regard to Property in Primitive Society. *American Anthropologist* New Series 37:244–256.
- (1937) 1939 The Culture Historical Method of Ethnology: The Scientific Approach to the Racial Question. New York: Fortuny. → First published as *Handbuch der Methode der kultur-historischen Ethnologie*.
- 1937–1942 Das Eigentum auf den ältesten Stufen der Menschheit. 3 vols. Münster in Westfalen (Germany): Aschendorff.
- 1944 Untersuchungen zur Methode der Ethnologie. *Anthropos* 35/36:898–965.
- 1952 Entstehung der Verwandtschaftssysteme und Heiratsregelungen. *Anthropos* 47:767–783.
- 1955 Das Mutterrecht. *Anthropos* Institut, Studia, Vol. 10. Vienna: Missionsdruckerei St. Gabriel. → Published posthumously.
- Wege der Kulturen: Gesammelte Aufsätze. *Anthropos* Institut, Studia, Vol. 20. St. Augustin bei Bonn (Germany): Verlag des Anthropos-Instituts, 1964.

## SUPPLEMENTARY BIBLIOGRAPHY

- BAUMANN, HERMANN 1958 P. Wilhelm Schmidt und Das Mutterrecht. *Anthropos* 53:212–228.
- BORNEMANN, FRITZ 1954 Verzeichnis der Schriften von P. W. Schmidt, S.V.D. (1868–1954). *Anthropos* 49:397–432.
- BORNEMANN, FRITZ 1958 P. W. Schmidts Studien über den Totemismus in Asien und Ozeanien. *Anthropos* 51:595–734.
- BORNEMANN, FRITZ 1958 P. W. Schmidts Studien über den Totemismus in Afrika. *Anthropos* 53:945–1003.



BURGMANN, ARNOLD (1954) 1966 P. W. Schmidt als Linguist. Volume 2, pages 287-328 in *Portraits of Linguists: A Biographical Source Book for the History of Western Linguistics, 1746-1963*. Edited by Thomas A. Sebeok. Bloomington: Indiana Univ. Press. → First published in Volume 49 of *Anthropos*.

Festschrift: *Publication d'hommage offerte au P. W. Schmidt*. Edited by Wilhelm Koppers. 1928 Vienna: Missionsdruckerei St. Gabriel.

HENNINGER, JOSEPH 1956 P. Wilhelm Schmidt, S.V.D. (1868-1954): Eine biographische Skizze. *Anthropos* 51: 19-60.

SEMAINE D'ETHNOLOGIE RELIGIEUSE *Comptes rendus*. → Proceedings of meetings held between 1912 and 1929, published in 5 volumes between 1913 and 1930.

## SCHMITT, CARL

Carl Schmitt, German political theorist, was born in 1888 in Plettenberg in Westphalia. He studied law and obtained his habilitation for public law at Strasbourg in 1916. His subsequent academic career was a successful one: he taught at Greifswald from 1921 to 1922; at Bonn from 1922 to 1928; at Berlin from 1928 until early in 1933, at Cologne during the year 1933, and again at Berlin, from the beginning of the academic year 1933/1934 until 1945.

Schmitt is an important representative of a school of thought in German public law which, after World War I, turned away from the German tradition of legal positivism and tried to broaden the scientific study of law and legal institutions by introducing into it historical, sociological, and political considerations. Accordingly, his publications cover a wide variety of subjects: the relationship between written law and judicial decision making, the romantic movement in political thinking during the early nineteenth century, and the history of dictatorship as a political idea and as an institution. He developed a systematic theory of constitutional law and made the history of European public law an object of philosophical investigation. He studied particular issues of constitutional, international, and martial law, and he engaged in polemics over the actual political situation in Germany.

Schmitt's participation in the antiliberal intellectual and political movement in Germany did not preclude his making significant contributions to political science. Observing the importance of arbitrary decisions both in the process of adjudication and in systematic jurisprudence, he established the concept of *Dezisionismus* to designate the fact that any legal order is ultimately based on sovereign decisions. These decisions are based not on positive norms but on the *ultima ratio* of emergency

decrees [see POLITICAL JUSTICE]. Schmitt traced this philosophy back to Hobbes's doctrine that *auctoritas non veritas facit legem* (laws are based on authority rather than truth), as well as to what he called the political theology of de Maistre, Bonald, Donoso Cortes, and other counterrevolutionary thinkers. Schmitt believed that the decisionistic method of looking for extreme and critical cases would lead to a new concrete way of thinking about the political order and political action (*konkretes Ordnungs- und Gestaltungsdenken*). However, the decisionistic method became more than an approach to political phenomena: Schmitt's preoccupation with the "extreme cases," as well as with the counterrevolutionary thinkers' pessimistic view of man, led him to convert the decisionistic method into a moral interpretation of politics and history that, in the turbulent situation in German politics between the two world wars, was to serve as a political ideology.

Schmitt defined politics as emerging from the friend-foe constellations that challenge the polity to distinguish between allies and enemies; friend-foe constellations arise in the situations in which civil war or war between states has to be considered an immediate possibility. All human groups, whether of economic, religious, or ethnic origin, may enter into extreme conflicts. The state, therefore, is valued because it is the institution which protects its citizens against the enemy within and without. Schmitt considered this defensive and repressive function of the state to be the precondition and the guarantee of all law and order in human affairs. He interpreted the modern history of continental Europe as a development from the absolute state of the eighteenth century, through the neutral state of the nineteenth, to the total state of the twentieth [see TOTALITARIANISM]. This total state would overcome the neutralization of political power and the cleavage between the state and society that had resulted from nineteenth-century liberalism. Liberalism and democracy were, in Schmitt's view, incompatible political principles, and the institutions of parliamentary government, once based on genuine political discussion and on the efficient functioning of public opinion, had, in twentieth-century mass society, lost their *raison d'être* and their legitimacy.

The Weimar Republic was, according to Schmitt, a distorted version of what he called a *parlamentarischer Gesetzgebungsstaat*, a distortion which he attributed to the waning of the nation's political homogeneity, which he considered to be caused in turn by the anarchical pluralism of incompatible political forces. Schmitt did not believe that Germany under the Weimar Republic could become

politically integrated. He therefore defended the National Socialists' seizure of power as a way of regaining the authority necessary to restore substantive justice. Thus he provided the ideological justification of a new German constitution. Based on the defense of national and racial homogeneity (*Artgleichheit*), and guided by the principles of leadership and loyalty, Germany should be rebuilt as a threefold totality comprising (1) a bureaucratic apparatus that included the judiciary, (2) a one-party political movement, and (3) a politically neutralized sphere of professional and communal administration (*Selbstverwaltung*).

In matters of international law, Schmitt criticized all global constructions of international legal order and defended a European Monroe Doctrine based on German hegemony. He distrusted concepts of international legal order because they could be used to justify imperialism and illegitimate imperialist intervention in the affairs of foreign states. He also deplored political universalism, because it produced a concept of warfare according to which both the enemy's state and its citizens could be prosecuted as criminals (the difference between the enemy and a criminal being blurred). [See INTERNATIONAL CRIMES.] As Germany approached defeat in World War II, Schmitt tried to discover what historical conditions had once enabled the European states to distinguish between an enemy and a criminal and to conduct war as an honorable duel between honorable adversaries. He believed he had found those conditions in a so-called *Nomos der Erde*, which originally had been shaped by the *jus publicum europaeum*. Schmitt believed that this *Nomos* could be found in a special geopolitical order, based on the division of the world into different areas of influence, dominated by the great European powers. These powers would recognize each other as independent sovereignties united by a common civilization of martial law.

Schmitt compared his own role under the Nazi regime to that of Benito Cereno in Herman Melville's story, who, after a mutiny on his slave ship had succeeded, served under the victors as a pirate, unable to disentangle himself from a situation in which protection was contingent on obedience. Schmitt's teaching, however, was often very near to the "mutinous" ideology which destroyed the Weimar Republic and helped the Nazi movement gain power. Yet by making the character of enmity, among citizens as well as nations, the crucial theme of systematic jurisprudence and political thinking, Schmitt created a new awareness of important problems.

JÜRGEN FIJALKOWSKI

[For the historical context of Schmitt's work, see CRISIS GOVERNMENT; DELEGATION OF POWERS; LEGITIMACY; NATIONAL SOCIALISM. Other related material may be found under JURISPRUDENCE; POLITICAL SCIENCE.]

#### WORKS BY SCHMITT

- 1912 *Gesetz und Urteil*. Berlin: Liebmann.
- (1919) 1925 *Politische Romantik*. 2d ed. Munich: Duncker & Humblot.
- (1921) 1928 *Die Diktatur: Von den Anfängen des modernen Souveränitätsgedankens bis zum proletarischen Klassenkampf*. 2d ed. Munich and Leipzig: Duncker & Humblot.
- (1922) 1934 *Politische Theologie*. 2d ed. Munich: Duncker & Humblot.
- (1923) 1961 *Die geistesgeschichtliche Lage des heutigen Parlamentarismus*. 3d ed. Berlin: Duncker & Humblot.
- 1926 *Die Kernfrage des Völkerbundes*. Berlin: Dümmler.
- (1927a) 1963 *Der Begriff des Politischen*. Berlin: Duncker & Humblot.
- 1927b *Volksentscheid und Volksbegehren*. Berlin: Gruyter.
- (1928) 1957 *Verfassungslehre*. 3d ed. Berlin: Duncker & Humblot.
- 1931 *Der Hüter der Verfassung*. Tübingen: Mohr.
- 1932 *Legalität und Legitimität*. Munich and Leipzig: Duncker & Humblot.
- 1933 *Staat, Bewegung, Volk*. Hamburg: Hanseatische Verlagsanstalt.
- 1934a *Staatsgefüge und Zusammenbruch des zweiten Reiches: Der Sieg des Bürgers über den Soldaten*. Hamburg: Hanseatische Verlagsanstalt.
- 1934b *Über die drei Arten des rechtswissenschaftlichen Denkens*. Hamburg: Hanseatische Verlagsanstalt.
- 1938a *Der Leviathan in der Staatslehre des Thomas Hobbes*. Hamburg: Hanseatische Verlagsanstalt.
- 1938b *Die Wendung zum diskriminierenden Kriegsbegriff*. Munich: Duncker & Humblot.
- 1939 *Völkerrechtliche Grossraumordnung, mit Interventionsverbot für raumfremde Mächte*. Berlin: Deutscher Rechtsverlag.
- 1940 *Positionen und Begriffe im Kampf mit Weimar-Genf-Versailles, 1923-1939*. Hamburg: Hanseatische Verlagsanstalt.
- (1942) 1944 *Land und Meer*. Leipzig: Reclam.
- 1950a *Donoso Cortés in gesamteuropäischer Interpretation*. Cologne: Greven.
- 1950b *Ex captivitate salus: Erfahrungen der Zeit 1945/47*. Cologne: Greven.
- 1950c *Der Nomos der Erde im Völkerrecht des Jus publicum Europaeum*. Cologne: Greven.
- 1963 *Theorie des Partisanen: Zwischenbemerkung zum Begriff des Politischen*. Berlin: Duncker & Humblot.
- Verfassungsrechtliche Aufsätze aus den Jahren 1924-1954: Materialien zu einer Verfassungslehre*. Berlin: Duncker & Humblot, 1958. → Essays originally published between 1924 and 1954, with commentaries by Schmitt on his own earlier articles.

#### SUPPLEMENTARY BIBLIOGRAPHY

- BARION, HANS; FORSTHOFF, E.; and WEBER, W. (editors) 1959 *Festschrift für Carl Schmitt zum 70. Geburtstag*. Berlin: Duncker & Humblot. → Contains a comprehensive bibliography of Schmitt's works and of works about Schmitt, compiled by Piet Tomissen.
- FIJALKOWSKI, JÜRGEN 1958 *Die Wendung zum Führerstaat: Ideologische Komponenten in der politischen Philosophie Carl Schmitts*. Cologne: Westdeutscher Verlag.



- FIJALKOWSKI, JÜRGEN 1965 Das politische Problem der Feindschaft. *Politische Vierteljahresschrift* 6:105-111.
- HOFMANN, HASSO 1964 *Legitimität gegen Legalität: Der Weg der politischen Philosophie Carl Schmitts*. Neuwied (Germany): Luchterhand.
- KROCKOW, CHRISTIAN 1958 *Die Entscheidung: Eine Untersuchung über Ernst Jünger, Carl Schmitt, Martin Heidegger*. Stuttgart: Enke.
- KRUPA, HANS 1937 *Carl Schmitts Theorie des "Politischen," mit einem Verzeichnis der Schriften Carl Schmitts*. Leipzig: Hirzel.
- SCHMIDT, HERMANN 1963 Der Nomosbegriff bei Carl Schmitt. *Staat* 2:81-108.
- SCHMITZ, MATHIAS 1965 *Die Freund-Feind-Theorie [in der politischen Philosophie] Carl Schmitts*. Cologne and Opladen: Westdeutscher Verlag.
- SCHNEIDER, PETER 1957 *Ausnahmestand und Norm: Eine Studie zur Rechtslehre von Carl Schmitt*. Stuttgart: Deutsche Verlags-Anstalt.
- WOHLGEMUTH, HEINRICH 1933 *Das Wesen des Politischen in der heutigen neoromantischen Staatslehre: Ein methodenkritischer Beitrag zu seiner Begriffsbildung*. Emmendingen: Dölter.

### SCHMOLLER, GUSTAV

Gustav Schmoller (1838-1917), German economist, came from a family of Württemberg civil servants, and he began his own career in the financial department of the Württemberg civil service. This experience gave him practical insights into the administration of a territory with a strong cameralist tradition and may have shaped his belief in the close interrelations between the economy and the government. At the University of Tübingen Schmoller studied *Staatswissenschaften*, a peculiarly German discipline derived from cameralism, which combines public finance, statistics, economics, administrative science, history, and even sociology. Later he became professor of *Staatswissenschaften* at the universities of Halle, 1864-1872; Strassburg, 1872-1882; and finally Berlin, 1882-1913.

In the early 1860s Schmoller defended the commercial treaty between France and the German Zollverein. Since he was supporting the point of view of the Prussian government against that of Württemberg, he ruined his career in Württemberg but gained favor with the Prussian authorities. In the Wilhelmian period, Schmoller became one of the most ardent Prussian patriots. He was appointed the official historian of Brandenburg and Prussia in 1887, became a member of the Prussian Staatsrat (state council) in 1884, and the representative of the University of Berlin in the Herrenhaus (upper house) in 1889. He was also a member of several academies (Berlin, Munich, St. Petersburg, Copenhagen, Vienna, and Rome) and

the recipient of honorary degrees from Breslau and Heidelberg.

As founder and leader of the Verein für Sozialpolitik, the association of German academic economists, and as editor or coeditor of several series of publications (*Staats- und sozialwissenschaftliche Forschungen*, from 1879; *Jahrbuch für Gesetzgebung, Verwaltung und Volkswirtschaft im Deutschen Reiche*—later known simply as *Schmollers Jahrbuch*—from 1881; *Forschungen zur brandenburgischen und preussischen Geschichte*, from 1888; and *Acta borussica*, from 1892) he became one of the first great organizers of research in the social sciences. Thus, he dominated the development of economics and of related social sciences in his country; for several decades, hardly a chair of economics was filled without his approval. He and his followers were called the "younger historical school" because their methodological approach was derived from, but also differed from, the approach of the "older historical school" of Roscher, Hildebrand, and Kries [see ECONOMIC THOUGHT]; they were also called the *Kathedersozialisten* (most nearly translated as academic socialists) because of their sociopolitical aims and were able to influence contemporary social legislation, since Schmoller was close to many high civil servants and to the crown.

In his political activities, Schmoller may be called a conservative social reformer. He was a royalist, favored strong government, and displayed the highest respect for the Prussian civil service; but he was also deeply preoccupied with the "social question," the improvement of the condition of the lower classes. He believed that a variety of means should be used to solve this problem: better education of the working classes; governmental regulation of working conditions; cooperatives; and changes in the law, in institutions, and in the attitudes of the ruling classes. He was hostile both to Manchesterism and to Marxism, but he was open to all kinds of reforms that would produce more "social justice," a concept central to his thinking.

Contributions to the social sciences. Schmoller's interests embraced history, applied statistics, economic and social institutions and behavior (including social psychology), social anthropology, geography, demography, the theory and practice of economic and social policy, and even ethics and philosophy, but he regarded himself primarily as an economist. His voluminous *Grundriss* (1900-1904), comparable in scope, though not in method, to Alfred Marshall's *Principles*, assembled pieces of all these social sciences into a great mosaic or panorama of the social world, past and present.

The only thing the work lacked was an analytical framework to hold the pieces together. Although Schmoller often claimed to have such a framework, he was basically eclectic.

Schmoller's rejection of analytical theory of the Smith-Ricardo type was of the utmost importance for his own thinking and for the development of economics in Germany under his "regime." He believed classical analysis to be useless or at best of minor importance, since it is applicable only to very small and artificially isolated fragments of the social reality that he wanted to perceive in all its many facets. It is not true, as many have asserted and as the term "historical school" suggests, that he wanted to replace analytical theory with purely historical research, but he was convinced that a genuine theory can only be the final outcome of an immense amount of descriptive work on past as well as present events, institutions, and structures. His conception of theory is the central defect in his thought and accounts for most of the weaknesses in his own works and in those inspired by him. It is difficult to decide whether he was led to this position by his inability to grasp the nature of theory or by his ethical disapproval of liberal economic doctrines, which he scorned as "business economics."

**Reception of Schmoller.** Schmoller's view of theory was attacked in 1883 by the Viennese professor Carl Menger, who wrote a treatise on method in the social sciences, *Problems of Economics and Sociology* (*Untersuchungen über die Methode der Socialwissenschaften und der politischen Oekonomie insbesondere*). Schmoller replied in his *Jahrbuch* (1883), and Menger renewed the attack in the pamphlet *Die Irrthümer des Historismus in der deutschen Nationalökonomie* (1884), "which fairly steamed with wrath and of course elicited rebuttal" (Schumpeter 1954, p. 814). This famous *Methodenstreit* dominated the economic and social sciences in Germany for at least two generations.

There can be little doubt that logically Menger was in a far better position. What he advocated as the appropriate methodology for economics was the use of the partial-equilibrium models that have since become the common method of theorists in most parts of the world. Nevertheless, Schmoller was not entirely wrong when he favored a much broader and more differentiated approach. He wanted to grasp all at once what has since become the object of many different branches of the social sciences. As a consequence, he touched on many important things, but often only superficially and seldom with clear insight about the implications of his observations and judgments. He was attacked

not only by theoretical economists but also by professional historians, such as Georg von Below (1904). Interestingly enough, both the theorists and the historians had basically the same objections to the Schmollerian type of social science: both missed a coherent and systematic approach. The issue, then, was not simply one of theory against history, but of how research in the social sciences ought best to be done. Although Schmoller often stated that his method alone was an "exact" one, even his friends and students agreed that he wrote without precision.

The very qualities of Schmoller's work that were criticized in the academic community won him a broad public audience. He took up many urgent issues of the day, examined them from different points of view, and came to general conclusions that many people could agree with. This made him an effective leader of public discussion and a popular, though not a brilliant, academic lecturer. He was at his best when dealing with those issues of social policy (*Sozialpolitik*) that urgently required compromise. In political debates he exhibited a considerable amount of personal commitment and civic courage, fighting against aggressive social conservatism (e.g. Treitschke 1875) as well as against the revolutionary position of Lassalle and Marx and their followers. In these debates he made plain his belief in the necessity of more social justice. However, he did not distinguish clearly between the value judgments he made in political debates and the axioms of his scientific thought. This led to another methodological attack on Schmoller, led by Max Weber and Werner Sombart, to eliminate value judgments from the social sciences (see Weber 1910; Sombart 1910).

**An appraisal of Schmoller.** Given his many weaknesses, one asks what made Schmoller the leader of German social scientists during the Wilhelminian period. Undoubtedly, others were stronger as theorists, as historians, or as sociologists. However, except for Max Weber, who was much younger, no one was quite as universal as Schmoller. Moreover, no one combined so many of the personal qualities of a leader: he was charming but dictatorial; confident in his views but able and willing to accept ideas; basically simple in his thinking but with a broad outlook; fully committed to sharply defined positions but prepared to compromise. Much of his influence appears to derive from his ability to express the leading ideas of his time. He swam with the stream of contemporary opinion, not against it, as is obvious in his views on foreign affairs, on which he had no ideas of his own but shared the nationalistic, imperialistic,



even chauvinistic outlook of the German elite. When the tide of sentiment and opinion turned, shortly after his death in 1917, his views were quickly outdated, and people began to wonder how he could ever have been the great leader he certainly was.

His influence on the development of the social sciences in Germany was, on the whole, rather unfortunate. He is mainly responsible for the neglect of economic theory in Germany between 1870 and 1920, and he can be made indirectly responsible for the extreme reaction that produced the subsequent neglect of economic history in twentieth-century Germany. In his attempt to construct a theory mainly by historical methods, he damaged both theory and history, and hampered their development in Germany for nearly a century. Outside Germany, his influence was never considerable, although American institutional economics was based partly on his ideas.

WOLFRAM FISCHER

#### WORKS BY SCHMOLLER

- 1870 *Zur Geschichte der deutschen Kleingewerbe im 19. Jahrhundert: Statistische und nationalökonomische Untersuchungen*. Halle: Waisenhaus.
- (1874-1875) 1875 *Über einige Grundfragen des Rechts und der Volkswirtschaft: Ein offenes Sendschreiben an Herrn Professor Dr. Heinrich von Treitschke*. Jena: Mauke. → First published in Volumes 23 and 24 of *Jahrbücher für Nationalökonomie und Statistik*.
- (1874-1897) 1904 *Über einige Grundfragen der Sozialpolitik und der Volkswirtschaftslehre*. 2d ed., enl. Berlin: Duncker & Humblot.
- 1879 *Die Strassburger Tucher- und Weberzunft; Urkunden und Darstellungen nebst Regesten und Glossar: Ein Beitrag zur Geschichte der deutschen Weberei und des deutschen Gewerberechts vom XIII.-XVII. Jahrhundert*. Strassburg: Trübner.
- 1883 *Zur Methodologie der Staats- und Sozialwissenschaften. Jahrbuch für Gesetzgebung, Verwaltung und Volkswirtschaft im Deutschen Reiche* 7:975-994.
- 1897 *Wechselnde Theorien und feststehende Wahrheiten im Gebiete der Staats- und Sozialwissenschaften und die heutige deutsche Volkswirtschaftslehre*. Berlin: Buxenstein.
- 1898 *Umriss und Untersuchungen zur Verfassungs-, Verwaltungs- und Wirtschaftsgeschichte besonders des preussischen Staates im 17. und 18. Jahrhundert*. Leipzig: Duncker & Humblot.
- (1900-1904) 1923 *Grundriss der allgemeinen Volkswirtschaftslehre*. 2d ed. 2 vols. Munich and Leipzig: Duncker & Humblot.
- 1913 *Charakterbilder*. 2 vols. Munich and Leipzig: Duncker & Humblot. → A collection of biographical essays, including those on Adam Smith, Friedrich List, Adolf Wagner, and Georg F. Knapp.
- 1918 *Die soziale Frage, Klassenbildung, Arbeiterfrage, Klassenkampf*. Edited by Luzie Schmoller. Munich and Leipzig: Duncker & Humblot. → Published posthumously. A revised and augmented edition of the sociological chapters contained in Schmoller's *Grundriss der allgemeinen Volkswirtschaftslehre*, 1900-1904.
- 1921 *Preussische Verfassungs-, Verwaltungs- und Finanzgeschichte*. Berlin: Tägliche Rundschau. → Published posthumously.
- 1922 *Deutsches Städtewesen in älterer Zeit*. Bonner Staatswissenschaftliche Untersuchungen, No. 3. Bonn and Leipzig: Schröder. → Published posthumously.
- Zur Literaturgeschichte der Staats- und Sozialwissenschaften*. Leipzig: Duncker & Humblot, 1888. → A collection of articles dated 1863-1883.
- Zur Social- und Gewerbepolitik der Gegenwart (Reden und Aufsätze)*. Leipzig: Duncker & Humblot, 1890. → A collection of articles dated 1872-1890.
- Zwanzig Jahre deutscher Politik (1897-1917): Aufsätze und Vorträge*. Munich and Leipzig: Duncker & Humblot, 1920. → Published posthumously.

#### SUPPLEMENTARY BIBLIOGRAPHY

- BELOW, GEORG VON 1904 *Zur Würdigung der historischen Schule der Nationalökonomie*. *Zeitschrift für Sozialwissenschaft* 7:145-185, 787-804.
- BELOW, GEORG VON 1924 *Zur Stellung G. Schmollers in der Geschichte der Nationalökonomie*. *Schmollers Jahrbuch für Gesetzgebung, Verwaltung und Volkswirtschaft im Deutschen Reiche* 48:315-319.
- BIERMANN, WILHELM E. (editor) 1922 *Briefwechsel zwischen Wilhelm Roscher und Gustav Schmoller*. Greifswald: Bamberg. → Letters dated 1861-1881.
- BRINKMANN, CARL 1937 *Gustav Schmoller und die Volkswirtschaftslehre*. Stuttgart: Kohlhammer.
- BRINKMANN, CARL 1953 *Gustav Schmoller*. Volume 9, pages 135-137 in *Handwörterbuch der Sozialwissenschaften*. Stuttgart: Fischer.
- EUCKEN, WALTER 1940 *Wissenschaft im Stile Schmollers*. *Weltwirtschaftliches Archiv* 52:468-506.
- GIDE, CHARLES; and RIST, CHARLES (1909) 1948 *A History of Economic Doctrines From the Time of the Physiocrats to the Present Day*. 2d English ed. Boston: Heath. → First published in French.
- HERBST, JURGEN 1965 *The German Historical School in American Scholarship: A Study in the Transfer of Culture*. Ithaca, N.Y.: Cornell Univ. Press.
- HERKNER, HEINRICH 1924 *Zur Stellung G. Schmollers in der Geschichte der Nationalökonomie*. *Schmollers Jahrbuch für Gesetzgebung, Verwaltung und Volkswirtschaft im Deutschen Reiche* 47:3-10.
- HINTZE, OTTO 1908 *Gustav Schmoller als Historiker*. Volume 4, pages 183-191 in Otto Hintze, *Historische und politische Aufsätze*. Berlin: Deutsche Bücherei.
- LANE, FREDERIC C. 1956 *Some Heirs of Gustav von Schmoller*. Pages 9-39 in *Architects and Craftsmen in History: Festschrift für Abbott Payson Usher*. Tübingen: Mohr.
- MEINECKE, FRIEDRICH 1933 *Drei Generationen deutscher Gelehrtenpolitik: Friedrich Theodor Visser.—Gustav Schmoller.—Max Weber*. Pages 136-164 in Friedrich Meinecke, *Staat und Persönlichkeit: Studien*. Berlin: Mittler.
- MEITZEL, C. 1925 *Gustav v. Schmoller*. Volume 7, pages 251-253 in *Handwörterbuch der Staatswissenschaften*. 4th ed. Jena: Fischer.
- MENGER, CARL (1883) 1963 *Problems of Economics and Sociology (Untersuchungen über die Methode der Sozialwissenschaften und der politischen Ökonomie insbesondere)*. Edited with an introduction by Louis Schneider. Urbana: Univ. of Illinois Press.

- MENGER, CARL 1884 *Die Irrthümer des Historismus in der deutschen Nationalökonomie*. Vienna: Holder. → The main attack against Schmoller and the historical school by an economic theorist and the culmination of the *Methodenstreit*.
- MITSCHERLICH, WALDEMAR 1936 Die ökonomischen Querschnitte Schmollers. Pages 129–149 in Waldemar Mitscherlich, *Die Lehre von den beweglichen und starren Begriffen erläutert an der Wirtschaftswissenschaft*. Stuttgart: Kohlhammer.
- RITZEL, GERHARD 1950 *Schmoller versus Menger: Eine Analyse des Methodenstreits im Hinblick auf den Historismus in der Nationalökonomie*. Frankfurt am Main: Enz & Rudolph.
- SALIN, EDGAR 1924 Zur Stellung G. Schmollers in der Geschichte der Nationalökonomie. *Schmollers Jahrbuch für Gesetzgebung, Verwaltung und Volkswirtschaft im Deutschen Reiche* 48:307–314.
- SCHUMPETER, JOSEPH A. 1926 Gustav v. Schmoller und die Probleme von heute. *Schmollers Jahrbuch für Gesetzgebung, Verwaltung und Volkswirtschaft im Deutschen Reiche* 50:337–388.
- SCHUMPETER, JOSEPH A. (1954) 1960 *History of Economic Analysis*. Edited by E. B. Schumpeter. New York: Oxford Univ. Press.
- SOMBART, WERNER 1910 Debatte [Über die Productivität der Volkswirtschaft]. Verein für Socialpolitik, *Schriften* 132:563–572.
- SPIETHOFF, ARTHUR 1924 Zum Abschluss. *Schmollers Jahrbuch für Gesetzgebung, Verwaltung und Volkswirtschaft im Deutschen Reiche* 48:321–324. → This article summarizes a series of assessments of Schmoller's work.
- SPIETHOFF, ARTHUR (editor) 1938 Gustav von Schmoller und die deutsche geschichtliche Volkswirtschaftslehre: Festgabe zur hundertsten Wiederkehr seines Geburtstages 24. Juni 1938. *Schmollers Jahrbuch für Gesetzgebung, Verwaltung und Volkswirtschaft im Deutschen Reiche* 62:385–757.
- TREITSCHKE, HEINRICH G. VON 1875 *Der Socialismus und seine Gönner: Nebst einem Sendschreiben an Gustav Schmoller*. Berlin: Reimer.
- WEBER, MAX 1910 Debatte [Über die Productivität der Volkswirtschaft]. Verein für Socialpolitik, *Schriften* 132:580–585, 603–607.

## SCHOOLS

See ACADEMIC FREEDOM; ADULT EDUCATION; EDUCATION; EDUCATIONAL PSYCHOLOGY; TEACHING; UNIVERSITIES.

## SCHRIEKE, BERTRAM

When the Dutch anthropologist Bertram Johannes Otto Schrieke (1890–1945) was studying at the University of Leiden, where he finished his studies with a doctor's degree in "languages and literature of the East Indian archipelago," sociology as a discipline was in its infancy in the Netherlands. It was taught only at the University of Amsterdam, where Schrieke attended the lectures of S. R. Steinmetz—through whom he was influ-

enced by the theories of Max Weber. At Leiden he was a pupil of Snouck Hurgronje, the eminent Islamologist and Arabist; of C. Van Vollenhoven, for adat law; and of other orientalist. His career was to be as diversified as his training.

Schrieke was not a theorist. Rather he took exception to existing theories as being too one-sided or too narrow. He held the view that "a culture forms an organic whole, which cannot be split up into different parts as if these components had no relation to each other," an approach not generally shared in the 1920s (*Indonesian Sociological Studies*, vol. 1, p. 232).

As demonstrated by his lifelong work in the social and economic aspects of the history of Indonesia and of Indonesian foreign contacts, Schrieke's forte lay in maintaining a critical attitude toward the generally accepted approaches to these subjects, which relied on archeology and colonial history.

In his doctoral thesis, *Het boek van Bonang* (1916), he analyzed the text of a manuscript, written in Javanese and ascribed to a legendary Muslim saint, which had been in the library of the University of Leiden since the beginning of the seventeenth century. His study of this manuscript, as well as of the early history of the Portuguese and Dutch trade contacts with Indonesia and of the published material on the voyages of Chinese and Arabs in the previous centuries, enabled him to revise the existing theory that the Islamization of the archipelago was the result of *pénétration pacifique*. He came to the conclusion that Islamization, hitherto considered to be primarily the product of trade relations, had been equally influenced by political conflicts and military struggles. In this work he acknowledged his debt to his mentor, G. P. Rouffaer, honorary member of the Royal Institute for Philology, Geography, and Ethnology of the Netherlands Indies at The Hague, whose own work of collecting and comparing source material was an inspiration to Schrieke.

Returning later to this problem in parts of a wide-ranging study, "Ruler and Realm in Early Java" (published posthumously, in an English translation only; 1957), Schrieke suggested that it is "impossible to understand the spread of Islam in the archipelago unless one takes into account the antagonism between the Muslim traders and the Portuguese," antagonism which caused the Indonesians to side with the Muslims.

During the 1920s, when Schrieke was acting as adviser on native and Arab affairs to the Netherlands Indies government, it was his intention to write a sociological study of the peoples of Sumatra



based on these ideas. He had written only the first, historical section of the "Prolegomena" to this study (see 1925) when he was called upon to conduct an inquiry into the so-called communist uprisings on the west coast of Sumatra in 1926. In order to explain the success of the communists' tactics, Schrieke incorporated in his report facts relating to the historical background and structure of Minangkabau society and dealt with the impact of the introduction of a money economy. He was thus able to point to the factors in this transitional period which had helped to foster the antigovernment attitude of the population. Here again his natural approach was to gather any relevant data, including the seemingly unimportant and the controversial. In 1955 the parts of his report that deal with the situation underlying the uprisings were published in an English translation, together with the above-mentioned "Prolegomena" and a number of his other publications (*Indonesian Sociological Studies*, vol. 1).

The Fourth Pacific Science Congress, convened at Batavia (now Jakarta) in 1929, afforded Schrieke the opportunity of editing a symposium entitled *The Effect of Western Influence on Native Civilisations in the Malay Archipelago*, to which his own contribution was "Native Society in the Transformation Period."

As director of the Netherlands Indies Department of Education from 1929 to 1933, Schrieke stressed the need for a sociological approach to educational problems. He defended the existing diversified educational system, adapted as it was to serve the needs of a heterogeneous population.

He did further field research in applied sociology at the invitation of the Julius Rosenwald Fund, which resulted in *Alien Americans* (1936). This was an "American Dilemma" *avant la lettre*: a one-man attempt to study the Negro problem in the United States and to interpret it against the background of other race relations there.

Apart from the influence he exerted in sociology and history when he was professor of sociology in the law school at Batavia, from 1924 to 1929, and later at the University of Amsterdam, from 1936 to 1945, Schrieke's ideas can best be traced in the work of J. C. van Leur (1934) and M. A. P. Meilink-Roelofs (1962). He fostered surveys of research on both Indonesia and the Netherlands Caribbean territories; he also encouraged such surveys in connection with the work of the International Refugee Colonization Society, established in 1938.

He also applied his organizational ability as museum director of the 150-year-old Batavian Society of Arts and Letters in 1923 and later, from

1938 to 1945, as director of the ethnological department of the Royal Colonial Institute (now Royal Tropical Institute), Amsterdam.

Schrieke died suddenly in London in September 1945 while attending a United Nations conference as a Netherlands government delegate on Indonesian affairs.

JOHANNA L. G. FELHOEN KRAAL

[See also *ASIAN SOCIETY*, article on SOUTHEAST ASIA; *COLONIALISM*; *ISLAM*; and the biography of SNOUCK HURGRONJE.]

#### WORKS BY SCHRIEKE

- 1916 *Het boek van Bonang*. Utrecht: Boer.
- 1925 *Prolegomena tot eene sociologische studie over de volken van Sumatra: Deel I* (A. Historisch gedeelte), *Schets van de politieke en economische machtsverschuivingen in den Indischen Archipel in de 17de eeuw* (Prolegomena to a Sociological Study of the Peoples of Sumatra: Part I [A. Historical Section], Outline of the Shifts in Political and Economic Power in the Indonesian Archipelago in the Seventeenth Century). *Tijdschrift voor taal-, land- en volkenkunde* 65:90-207. → For a partial translation, see pages 1-82 of *Indonesian Sociological Studies: Selected Writings of B. Schrieke*.
- (1926) 1928 *The Evolution of Culture in the Pacific in Relation to the Theories of the "Kultur-Historische" and the "Manchester" Schools of Social Anthropology*. Volume 2, pages 2423-2441 in *Pacific Science Congress, Third, Tokyo, Proceedings*. Tokyo: National Research Council of Japan.
- 1929 *Native Society in the Transformation Period*. Pages 237-247 in Bertram Schrieke (editor), *The Effect of Western Influence on Native Civilisations in the Malay Archipelago*. Batavia (Java): Kolff.
- 1936 *Alien Americans: A Study of Race Relations*. New York: Viking.
- 1938 *The Educational System in the Netherlands Indies*. *Netherlands Indies: Bulletin of the Colonial Institute*, Amsterdam 2: 14-24.
- 1948 SCHRIEKE, BERTRAM (editor) *Report of the Scientific Work Done in the Netherlands on Behalf of the Dutch Overseas Territories During the Period Between Approximately 1918 and 1943*. Amsterdam: North Holland Publishing. → Published posthumously.
- 1957 *Ruler and Realm in Early Java*. Volume 2, pages 1-267 in *Indonesian Sociological Studies: Selected Writings of B. Schrieke*. The Hague: Van Hoeve. → Published posthumously from Schrieke's Dutch manuscript.
- Indonesian Sociological Studies: Selected Writings of B. Schrieke*. 2 vols. The Hague: Van Hoeve, 1955-1957. → Contains parts of the 1925 *Prolegomena to a Sociological Study of the Peoples of Sumatra*, parts of the 1926 report on uprisings on the west coast of Sumatra, and a number of later writings, some previously unpublished.

#### SUPPLEMENTARY BIBLIOGRAPHY

Foreword by the Editors of "Selected Studies on Indonesia by Dutch Scholars." 1957 Volume 2, pages v-vii in Bertram J. O. Schrieke, *Indonesian Sociological Studies: Selected Writings of B. Schrieke*. The Hague: Van Hoeve.

- LEUR, JACOB C. VAN (1934) 1955 *On Early Asian Trade*. Pages 1-44 in Jacob C. van Leur, *Indonesian Trade and Society: Essays in Asian Social and Economic History*. The Hague: Van Hoeve. → First published in Dutch.
- LINDGREN, ETHEL J. 1948 Bertram Johannes Otto Schrieke: 1890-1945. *Man* 48:113-117. → Includes a bibliography on pages 116-117, prepared by Lindgren and her associates.
- MEILINK-ROELOFSZ, MARIE A. P. 1962 *Asian Trade and European Influence in the Indonesian Archipelago Between 1500 and About 1630*. The Hague: Nijhoff.

## SCHULTZ, HENRY

Henry Schultz (1893-1938), American econometrician, was born in Russian Poland. After attending public schools in New York City and the College of the City of New York, he began his graduate studies at Columbia University in 1916. At Columbia he responded enthusiastically to the influence of Henry L. Moore, but his studies were interrupted by military service in World War I. After the war an army scholarship enabled him to spend the spring and summer terms of 1919 at the London School of Economics and Political Science and at the Galton Laboratory of University College. In London he attended lectures on statistics by A. L. Bowley and Karl Pearson and on economic theory by Edwin Cannan. Once back in the United States, he conducted statistical and economic investigations for, successively, the War Trade Board, the Bureau of the Census, the Bureau of Efficiency, the Institute of Economics, and the Children's Bureau of the United States Department of Labor. His Ph.D. thesis, "The Statistical Law of Demand as Illustrated by the Demand for Sugar," was published in 1925, and in 1926 he joined the department of economics at the University of Chicago, where he remained until his death.

It was Moore's early investigations into the statistical approximation of Marshallian demand curves (1914; 1917) that had a determining influence on Schultz's career. Under Moore's inspiration, Schultz made intensive studies of the works of those economists—Cournot, Walras, and Pareto—who had made the most explicit and sophisticated use of mathematics in formulating economic theory. Schultz also studied probability theory and statistics in considerable depth. His years in various statistical and economic research agencies in Washington, 1920 to 1926, no doubt contributed to his sophistication in the use and interpretation of published economic data.

When Moore ceased writing in 1929, Schultz was almost the only American economist who com-

bined a mastery of the theoretical work of Walras and Pareto with an intense interest in empirical studies of demand. He was fully abreast of the revolution in value theory associated with the names of J. R. Hicks and R. G. D. Allen (1934); they also drew their inspiration from Walras and Pareto. Keynesian macroeconomic theory, however, was not of direct significance to Schultz's work, and the lengthy bibliography of his most important book, *The Theory and Measurement of Demand* (1938, pp. 779-803), contains no reference to Keynes.

Schultz published a number of short notes on regression theory and methods and one major article (1930) on the derivation of standard errors of forecast for curves of various types. In *Statistical Laws of Demand and Supply* (1928), he thoroughly explored the literature on weighted regression—the estimation of a single line or plane of best fit when all variables are subject to error—and was probably the first to apply such methods to economic data. This book presented supply-and-demand curves for sugar, an internationally traded commodity, and led him to write three articles on methods of measuring the effectiveness of a tariff. His contributions to statistical theory must be regarded as by-products of his intense concern with statistical demand analysis, and the work on the tariff was similarly incidental to his main concern. Schultz also published a number of expository articles on the more general contributions of mathematics and statistics to economic research and in book reviews called attention to significant translations or interpretations of Walras and Pareto and to major works on mathematical economics and statistics.

Schultz's lifework is essentially summed up in his monumental 817-page work *The Theory and Measurement of Demand* (1938). Chapters 1, 18, and 19 together provide an excellent statement of the theory of consumer demand, incorporating the Hicks-Allen and Slutsky extensions of Pareto's basic formulation; Chapters 2 and 3 deal with the logical and theoretical validity of various proposed methods for deriving demand curves from time series and from family-budget data; Chapter 4 contains a statement of the assumptions, methods, and procedures to be used in the empirical section. The empirical section, Chapters 5-17, presents statistical demand analyses for ten major crops along with summary comparisons and interpretations of the results for the various commodities, time periods, and functional forms. Further empirical studies are included in Chapters 18 and 19, on related demands. A fifty-page appendix on ele-



sions of Book v of Marshall's *Principles* and other theoretical topics. At Harvard he taught one of the first modern courses in pure theory.

His lectures were never dull; often they were brilliant. But what distinguished Schumpeter's lectures and seminars most was neither their brilliance nor their organization (which occasionally left something to be desired); rather, it was his persistent effort on all occasions, formal or informal, to make it possible for students to work out their own ideas. In nine years of attending his lectures and seminars in Bonn and at Harvard, I remember only two occasions when he talked about his own work and then only at the prodding of his students.

He spent much time with his students outside the classroom. In Bonn he frequently went on walks with them, discoursing, interrupting his walk suddenly to jot down an idea. At Harvard he regularly met his students socially. He also circulated a sheet of paper during each lecture on which six students signed up to see him. He was not the kind of teacher who provided his students with clear-cut answers, to the despair of the less-gifted students. But for others this method had a delayed effect that years later would make them sit up and exclaim, "So this is what Schumpeter meant!"

### The Schumpeterian system

Because of Schumpeter's many-sided gifts and the variety of fields to which he contributed, it is difficult to trace his intellectual ancestry. His teachers at the University of Vienna included Friedrich von Wieser, E. von Phillipovich, and E. von Böhm-Bawerk. But among all economists, he had the greatest esteem for Léon Walras, and he thought very highly of English and American economists, particularly F. Y. Edgeworth and John Bates Clark. The similarity of Schumpeter's thought to Marx's has frequently been noted, although Marx's influence on Schumpeter may have been overestimated because of Schumpeter's ability to admire and understand Marx even when he disagreed with him.

Schumpeter was more aware than many other theoretical economists, the most notable exception being Marx, that economic reality is only a part of total reality. The division of the economic from the noneconomic was for him purely a matter of scientific convenience. He contributed significantly to the understanding of noneconomic, particularly sociological, phenomena, as in his synthesis, *Capitalism, Socialism, and Democracy* (1942). If his work lacks the mathematical precision he urged

upon others, this was not the result of unclear thinking but of the complexity (one might almost say the messiness) of the problems he tackled.

Any brief characterization of an elaborate and subtle system is necessarily something of a caricature. Bearing this limitation in mind, Schumpeter's vision of the economic process might be summarized as follows:

There is only one single historical reality, which consists of unique "events." To understand this reality one must always be aware of its quality of uniqueness. But one must also realize that events are related to each other, not just simultaneously but over time. Simultaneous relationships are the subject of equilibrium theory; intertemporal relationships are the subject of dynamic theory. Both equilibrium theory and dynamic theory are necessary to explain any event at any given moment of time.

**Statics and dynamics.** The essential difference between statics and dynamics—a dichotomy that Schumpeter distinguished from the dichotomy between stationary and evolutionary—is not fully expressed by stating that dynamics involves time and statics does not. For Schumpeter static, or equilibrium, theory deals with processes of adaptation that are found in reality: he believed he could identify particular periods in history in which the economic system was "in the neighborhood of equilibrium." His equilibrium, however, has no normative content; it is not in any sense an ideal state of affairs, a golden age.

It follows, therefore, that the dynamics that concerned Schumpeter was *not* the time sequence of adjustment to some change that had somehow occurred. Rather, he sought to discover what causes the movement away from equilibrium, what is responsible for the destruction of equilibrium. For reasons of scientific convenience, he centered his attention on those disruptions of equilibrium that arise inside the economic system, although he believed that equilibrium can also be disrupted by forces outside the economy, such as wars or natural catastrophes. (While for Marx, and certainly for Lenin, wars arise from the inner contradictions of capitalism, Schumpeter stressed the essentially pacific and constructive nature of capitalism and its evolution.)

**Business cycles.** The distinction between statics and dynamics is fundamental to Schumpeter's ideas on the business cycle. It explains not only why a recession, being an adaptive phenomenon, must follow a boom, but also why a boom requires a different explanation from a recession. It shows

both why historical data are needed to arrive at an explanation and how such data are fitted into a general framework and thereby become meaningful. It explains in what sense business cycles are essential to progress: business cycles occur because equilibrium is destroyed by innovations, and the economy must adapt to the destruction before there can be any further innovation. Finally, the distinction accounts for the institutional requirements that must be met if an economy is to remain dynamic. The essential institutional arrangements are two: first, scope for the operation of innovators, the Schumpeterian entrepreneurs; and second, the possibility for innovators to break into the circular flow of the economy, this being achieved through the creation of credit. Credit creation permits the entrepreneur to take resources out of the income stream for productive purposes *before* he has contributed anything to the income stream.

Schumpeter distinguished three types of business cycles, each of whose length depends on the disturbance that brought it about. He named the cycles for pioneers in business cycle theory. The shortest, the Kitchin cycle, is identified with inventory accumulation and decumulation lasting about three years. The medium cycle, of about 8–11 years, usually referred to as *the* business cycle, he named for Clément Juglar, and he related it to individual innovations, such as new textile machines, dynamos, electric motors, radios, and refrigerators. The long cycle, named for Kondratieff, is caused by the appearance of major innovations, such as railroads or electrification, which cannot be carried through in one “juglar.” Thus, long-run trends become part of cyclical phenomena and, indeed, can be understood as the result of shorter-run changes.

Schumpeter believed that each cycle consists of four phases. He characterized the recovery and recession phases as adaptive processes, while the upswing and depression phases represent movements away from the neighborhood of equilibrium, caused, respectively, by innovation and by such extrinsic and inessential phenomena as speculation, panics, or poor economic policies. The upswing and recession phases are essential to the capitalist process, but the depression and recovery phases are not. Since the acute social and personal hardships occur in the inessential phases, policy can eliminate them.

**Development of Schumpeter's system.** Beginning in 1908, Schumpeter published a series of books in which he progressively developed these economic theories. His first major book was *Das*

*Wesen und der Hauptinhalt der theoretischen Nationalökonomie* (1908), which was probably conceived during his days as a graduate student and written in Cairo. It was a brilliant statement of general equilibrium theory, making clear both what static theory could explain and what it could not. This was followed in 1912 by *The Theory of Economic Development*, which contained the first important vision of the “Schumpeterian system”—a term he would have abhorred. The book dealt primarily with dynamics—the entrepreneurial function, the role of the banking system, and the creation of credit. Finally, in *Business Cycles* (1939), he fleshed out the bones of theory with historical material. Moreover, *Business Cycles* contains numerous partial mathematical models.

**Assessment of the system.** Schumpeter's detailed discussion of historical data in the context of his theory has, of course, led to criticism of his interpretation of particular facts. Also, it has been suggested that his emphasis on the innovating entrepreneur tends to limit his theory to the past. Schumpeter himself felt that reality fitted his three-cycle schema better than his theory warranted, since nothing in his theory required or even suggested any particular regularity of timing. He also felt that he had accomplished his task if he had adequately explained capitalism from the eighteenth century to the early twentieth century, and he apparently agreed that in the future trust-dominated capitalism and planned socialism might make his theory obsolete.

Many years later Schumpeter wrote that he was “trying to construct a theoretic model of the process of economic change in time, or perhaps more clearly, to answer the question how the economic system generates the force which incessantly transforms it” (*Essays*, pp. 158–159). Because of this, and despite his own skepticism about the future relevance of his theory, Schumpeter's approach retains its usefulness in underdeveloped as well as in developed countries, and in socialist as well as in market economies. In order for there to be economic development, the Schumpeterian problem of when to introduce a change, when to introduce an investment, must be solved. There is nothing automatic about the introduction of new investments, and planners, in both market and socialist economies, must think like Schumpeterian entrepreneurs if their plans are to lead to self-sustaining growth. If a theory is understood to be a collection of formulas that can be applied with a minimum of analysis and thought, then Schumpeter's theory is indeed obsolete. But if it is understood to be a



view of the economy that enables a policy maker to ask pertinent questions and develop appropriate programs of action, then the relevance of Schumpeter's thought to any changing economy cannot be doubted.

### Other economic writings

Schumpeter's contribution to the history of economics, like his contribution to economic theory, developed over many years. The posthumously published *History of Economic Analysis* (1954) is an elaboration of his much earlier *Economic Doctrine and Method* (1914). Both are histories of economic theories, which Schumpeter believed develop in the same manner as do theories in the natural sciences. In other words, economic theories, unlike philosophies, are not self-contained and develop over time in a cumulative fashion. The *History of Economic Analysis* was published from an unfinished manuscript. It is, by any standard, an outstanding intellectual achievement. While it met with detailed criticism (for example, Schumpeter was charged with having been less than fair to the neoclassical economists), it is, of course, impossible to say to what extent these criticisms would have been avoided if Schumpeter had lived to finish the work.

Schumpeter also wrote many articles on monetary theory, relating monetary theory to his general view of development. As early as 1917, when it was not common to do so, he discussed money in relation to national income and economic policy. While he was writing a book on monetary theory, Keynes's *Treatise on Money* appeared. Schumpeter admired the *Treatise* and insisted that all his students read it, but he felt that its substance required a complete rewriting of his own work.

### Sociological writings

Schumpeter considered from a sociological point of view a variety of phenomena that others had interpreted in mainly economic terms—imperialism, social classes, taxation, socialism, and even the entrepreneur. He attempted to relate economic phenomena to the larger social context. Thus, for example, he rejected the view that economic explanations can adequately account for the phenomenon of imperialism, which is an "atavism" of social structure and of emotional habits of an individual-psychological nature, and he stressed that any explanation of the entrepreneur must include the social and political conditions necessary for him to fulfill his functions.

Schumpeter brought together his sociological and economic analyses primarily in *Capitalism,*

*Socialism, and Democracy* (1942), a work remarkable for its style as well as for its vision. Essential to his argument is his analysis of Marxian doctrine. Schumpeter did not think that capitalism shows the inherent contradictions that Marx claimed would lead to the immiserization of the masses and to revolution; nor does capitalism per se lead to imperialistic adventures or to wars. But he sympathized with the Marxian idea that the achievements of capitalism, which Marx himself had praised so highly, have their own inherent logic: because capitalism has a historic function to perform, it will come to an end as the function disappears. Socialism is possible, perhaps even inevitable, but what Schumpeter meant by socialism is different from what is commonly understood by the term.

Socialism is defined by Schumpeter as "an institutional pattern in which the control over means of production and over production itself is vested with a central authority—or, as we may say, in which, as a matter of principle, the economic affairs of society belong to the public and not to the private sphere" (1942, p. 167). This is quite consistent with the fact that "some freedom of action must be left, and almost any amount of freedom might be left, to the 'men on the spot,' say, the managers of the individual industries or plants" (p. 168). It is noteworthy that the definition does not necessarily include ownership of the means of production.

Schumpeter made short shrift of the numerous reasons that have been adduced to prove the approaching collapse of capitalism. He argued that capitalism cannot be even approximately identified with perfect competition and, therefore, to criticize it for monopolistic practices or for temporary recessions misses the point. "Capitalism . . . is by nature a form or method of economic change and not only never is but never can be stationary" (1942, p. 82), and most criticism overlooks this essential fact.

The Achilles' heel of capitalism lies elsewhere, namely in its achievements. The very success of capitalism will bring about its doom for two reasons: as societies grow richer, and production and productivity increase spectacularly, the need for further growth and for the specific contribution of the entrepreneur will disappear. Administrators can deal with what will become essentially routine problems, and the desiderata of economic life will shift. At the same time, the successes of capitalism will lead to an increasingly hostile intellectual environment, to a *trahison des clercs*, in which entrepreneurs will find it increasingly difficult to fulfill

their essential function. Schumpeter was also prepared to consider that the process of innovation might itself become more and more routine. He foresaw the eventual triumph of socialism, but he thought its chances for success and indeed its relevance negligible in precisely those situations where most people believed (and still believe) it necessary—to make poor economies grow.

Economically, capitalism has performed miracles, and in another fifty to one hundred years it will have created such a powerful productive machine as to eliminate poverty. (Indeed, Schumpeter's own projections—in Chapter 5—have already been far surpassed.) Culturally, capitalism has brought about the spread of a rational and pacific view of the world: "Modern mathematico-experimental science developed . . . not only along with the social process usually referred to as the Rise of Capitalism, but also outside . . . of scholastic thought and in the face of its contemptuous hostility" (1942, p. 124). The same is true of painting and literature.

Schumpeter also discussed the problem of the transition to socialism. The socialism that can work is one that comes "in the fullness of time," when the problem of poverty has been solved by the preceding creation of a powerful productive economy. Schumpeter believed that he and Marx were basically in agreement on this point:

Evolution was for him [Marx] the parent of socialism. He was much too strongly imbued with a sense of the inherent logic of things social to believe that revolution can replace any part of the work of evolution. . . . The Marxian revolution therefore differs entirely, in nature and in function, from the revolutions both of the bourgeois radical and of the socialist conspirator. It is essentially revolution in the fullness of time. . . . To say that Marx, stripped of phrases, admits of interpretation in a conservative sense is only saying that he can be taken seriously. (1942, p. 58)

If socialism is built on a powerful productive economy, then the problems of transition will be negligible; if this condition does not obtain, these problems of transition may jeopardize the success of socialism. Schumpeter also considered the rationality of socialist decision making, first discussed by Enrico Barone in 1905, and the human problems involved.

According to Schumpeter, socialism can be run on democratic lines. His definition of democracy may not, however, command general assent—an "institutional arrangement for arriving at political decisions in which individuals acquire the power to decide by means of a competitive struggle for the people's vote" (p. 269). What will actually

happen will depend on many accidents as well as on the inherent logic of social (including economic) evolution: "... there is little reason to believe that . . . socialism will mean the advent of the civilization of which orthodox socialists dream. It is much more likely to present fascist features. That would be a strange answer to Marx's prayer. But history sometimes indulges in jokes of questionable taste" (p. 375).

It is difficult to present Schumpeter's thought in a brief form: its complexity is such that the many pages Schumpeter himself devoted to it are entirely necessary. This complexity might have prevented the development of a Schumpeter school, even if he himself had not actively discouraged its formation. Moreover, it is not easy to locate Schumpeter in a history of economic analysis: although he was an Austrian by birth and training, he was not an "Austrian" economist. He belongs to that recent trend in economics that is characterized by an absence of schools of economic thought, though not of good economists. The economists Schumpeter himself trained will never put reality into the strait-jacket of any particular theory; they will be able to use any logical theory while remaining conscious of the variety and uniqueness of the real social and economic problems.

WOLFGANG F. STOLPER

[See also BUSINESS CYCLES; CAPITALISM; ECONOMIC THOUGHT, article on THE AUSTRIAN SCHOOL; ENTREPRENEURSHIP; and the biographies of BÖHM-BAWERK; CLARK, JOHN BATES; EDGEWORTH; MARX; WALRAS; WIESER.]

#### WORKS BY SCHUMPETER

- 1908 *Das Wesen und der Hauptinhalt der theoretischen Nationalökonomie*. Leipzig: Duncker & Humblot.
- (1912) 1934 *The Theory of Economic Development: An Inquiry Into Profits, Capital, Credit, Interest, and the Business Cycle*. Harvard Economic Studies, Vol. 46. Cambridge, Mass.: Harvard Univ. Press. → First published as *Theorie der wirtschaftlichen Entwicklung*.
- (1914) 1954 *Economic Doctrine and Method: An Historical Sketch*. London: Allen & Unwin. → First published as *Epochen der Dogmen- und Methodengeschichte*.
- (1918) 1954 *The Crisis of the Tax State. International Economic Papers 4:5-38*. → First published as *Die Krise des Steuerstaats*.
- (1919-1927) 1951 *Imperialism and Social Classes*. New York: Kelley. → First published as "Zur Soziologie der Imperialismen" (1919) and "Die sozialen Klassen im ethnisch homogenen Milieu" (1927). A paperback edition was published in 1955 by Meridian.
- 1939 *Business Cycles: A Theoretical, Historical, and Statistical Analysis of the Capitalist Process*. 2 vols. New York and London: McGraw-Hill. → An abridged version was published in 1964.



- (1942) 1950 *Capitalism, Socialism, and Democracy*. 3d ed. New York: Harper; London: Allen & Unwin. → A paperback edition was published by Harper in 1962.
- (1954) 1964 *History of Economic Analysis*. Edited by E. B. Schumpeter. New York: Oxford Univ. Press. → Published posthumously.
- Aufsätze zur ökonomischen Theorie*. Edited by Arthur Spiethoff and Erich Schneider. Tübingen (Germany): Mohr, 1952.
- Aufsätze zur Soziologie*. Edited by Arthur Spiethoff and Erich Schneider. Tübingen (Germany): Mohr, 1953.
- Dogmenhistorische und biographische Aufsätze*. Edited by Arthur Spiethoff and Erich Schneider. Tübingen (Germany): Mohr, 1954.
- Essays*. Edited by Richard V. Clemence. Cambridge, Mass.: Addison-Wesley, 1951. → Contains essays first published between 1909 and 1950.
- Ten Great Economists: From Marx to Keynes*. New York: Oxford Univ. Press, 1951. → Contains essays first published between 1910 and 1950. A paperback edition was published in 1965.

## SUPPLEMENTARY BIBLIOGRAPHY

- HABERLER, GOTTFRIED 1950 Joseph Alois Schumpeter: 1883-1950. *Quarterly Journal of Economics* 64:333-372.
- HABERLER, GOTTFRIED 1951 J. A. Schumpeter, 1883-1950. Pages 24-47 in S. E. Harris (editor), *Schumpeter: Social Scientist*. Cambridge, Mass.: Harvard Univ. Press.
- HARRIS, SEYMOUR E. (editor) 1951 *Schumpeter: Social Scientist*. Cambridge, Mass.: Harvard Univ. Press.
- KESSELER, MARTIN 1961 The Synthetic Vision of Joseph Schumpeter. *Review of Politics* 23:334-355.
- SCHUMPETER, ELIZABETH B. 1950 Bibliography of the Writings of Joseph A. Schumpeter. *Quarterly Journal of Economics* 64:373-384.
- SMITHIES, A. 1951 Memorial: J. A. Schumpeter, 1883-1950. Pages 11-23 in S. E. Harris (editor), *Schumpeter: Social Scientist*. Cambridge, Mass.: Harvard Univ. Press.
- SPIETHOFF, ARTHUR 1949 Josef Schumpeter: In Memoriam. *Kyklos* 3:289-293. → Dated 1949, but published after Schumpeter's death in 1950.
- WRIGHT, DAVID MCCORD 1950 Schumpeter and Keynes. *Weltwirtschaftliches Archiv* 65, no. 2:185-196.

## SCHUTZ, ALFRED

Alfred Schutz (1899-1959) was born in Austria, studied at the University of Vienna, and from 1939 until his death lived in New York City, where he was professor of philosophy and sociology at the New School for Social Research. His principal philosophical work consisted of the application of Edmund Husserl's phenomenology to the problems of social reality, while his major methodological contribution was an attempt to relate phenomenological concepts to the sociology of Max Weber. He believed that the cardinal problem for the social sciences is the study of the world of daily life, of the common-sense reality that each individual shares with his fellow men in a taken-for-granted

manner, and both his major intellectual efforts focused on this problem. As Schutz saw it, the task of the phenomenological philosopher concerned with social reality is to uncover, describe, and analyze the essential features of this mundane world; and consequently all of his writings, beginning with *The Phenomenology of the Social World* (1932), take the reality of everyday life as a point of departure and as a subject for detailed examination.

For Schutz, the world of common-sense reality rests on certain basic epistemological presuppositions: that the mundane world exists, that man and his fellows exist in that world, and that man can communicate with Others. In daily life, men take for granted the reality of their experience: they assume, naively, that they share the same world with all other "normal" perceivers; and they assume that were they to change places with Others, they would see the world in essentially the same way the Others do.

Upon analysis, it is found that these fundamental presuppositions imply a number of structural characteristics. (1) The "biographical situation" of the individual depends on his particular location in the social world as he himself defines it and as he interprets that location in terms of his previous experiences. The biographical situation is unique to the individual, although it includes typical features he does share with Others. (2) Common-sense man possesses a "stock of knowledge at hand." It includes elements handed down to him by Others, parents and teachers among them, and it is in a typically organized form. This knowledge is accepted without question as generally valid, though it is always potentially questionable. (3) The spatial coordinates of "here and there" are essential features of the social matrix by which the individual locates himself. The individual's own body is the reference point in terms of which he locates himself as "here" and the Other as "there." (4) The "alter ego" is given to the individual as a being like himself in a mundane world. Moreover, the Other exists simultaneously with the self, his thought is grasped in a vivid present, and alter ego's speech and the individual's listening coexist in a relationship that is the essence of intersubjectivity. (5) Ego's world, as an intersubjective social reality, includes alter egos with different temporal and spatial characteristics: predecessors, or those who lived before him and are known to him through report, contemporaries, who share the same temporal world; consociates, or those who are alive at the same time and who also share ego's spatial segment of the world through face-to-

face relationships; successors, who will live after he dies, including those who will be born only after he is dead. The social world is as much constituted by consciousness of predecessors and successors as it is by contemporaries and consociates. To explain the structural relationships that hold among all of these aspects of social order, a theory of social action is necessary.

Schutz distinguished between "action" and "act." Action is understood as purposive conduct projected by the actor; an act is defined as accomplished action. Action has a subjective form as well as an objective form: purposively *refraining* from doing something is no less action than overtly behaving in one way or another. What is of decisive importance to Schutz's conception of action is his view that action is grounded in the interpretative consciousness of the actor. It is at this point that Schutz turned to Weber, in particular to his conception of interpretative understanding (*Verstehen*) and to his postulate of the subjective interpretation of meaning. To understand social action is to grasp the meaning which the actor gives to or bestows upon his action, to comprehend what that action means to *him*. The "subjective" interpretation of meaning signified for Weber what the actor *means* by his action, not, as Schutz stressed, some private idiosyncratic and unverifiable domain.

In Schutz's adaptation of Weber, subjective interpretation of meaning appears primarily as a fundamental typification of the common-sense world. All men in mundane life, then, practice *Verstehen*, but there is a distinction between *Verstehen* (1) as an experiential form of common-sense knowledge, (2) as an epistemological problem, and (3) as a distinctive method of the social sciences. Schutz believed that it was the great merit of such men as Durkheim, Pareto, Marshall, and Veblen, and above all of Weber, to have developed the technique that "consists in replacing the human beings which the social scientist observes as an actor on the social stage by puppets created by himself, in other words, in constructing ideal types of actors" (*Collected Papers*, vol. 2, p. 17). The social scientist, according to Schutz, proceeds as follows:

[He] observes certain events within the social world as caused by human activity and he begins to establish a type of such events. Afterwards he coordinates with these typical acts typical because motives and in-order-to motives which he assumes as invariable in the mind of an imaginary actor. Thus he constructs a personal ideal type, which means the model of an actor whom he imagines gifted with a consciousness. . . . The social scientist places these constructed types in

a setting which contains all the elements of the situation in the social world relevant for the performance of the typical act under inquiry. Moreover, he associates with him other personal ideal types with motives apt to provoke typical reactions to the first ideal type's typical act. So he arrives at a model of the social world, or better at a reconstruction of it. (*ibid.*, pp. 17-18)

Action is never an isolated phenomenon; it has its phenomenological "horizons" of relevance and relatedness to social reality, and it has its "motivation." As already noted, typical acts have two kinds of motives, "in-order-to" motives and "because" motives. The first are explained in terms of the actor's goals and ends; the second, in terms of his background and disposition. The in-order-to motive is "the future state of affairs to be realized by the projected action," and the project itself is determined by the because motive, which in turn is dominated by the past tense. Other people's acts can be fully understood only insofar as their because and in-order-to motives are known, but since this is impossible, it must suffice to know their typical motives, including their reference to typical situations, typical ends, typical means, etc. (*ibid.*, pp. 11-12).

The temporal disequilibrium between the two orders of motive suggests a larger problem in the conception of the ego. According to Schutz, the ego cannot seize its own immediacy; it can capture itself only as an object of a reflexive act. Schutz here came close to George H. Mead's distinction between the "I" and "me" aspects of the self. He pointed, however, to a larger implication of the distinction, for man is understood as a being who presents himself to Others, takes his place in the social world, and, finally, knows himself only in a partial and fragmentary way. Such fragmentation is itself a primordial typification of reality.

Schutz's philosophical position has significance for the choice of models to be used in the social sciences, and it suggests, in particular, that the use of natural science models is not satisfactory. He believed that the natural and social sciences differ qualitatively: the natural sciences investigate objects that are constructs of the first degree, i.e., objects in the field of the scientific observer; the social sciences are properly concerned with second-degree constructs, i.e., objects that not only are themselves in a world but that also *have* a world. The human beings who are the central concern of the social scientist are interpreters of their own lives and actions. The task of the social scientist is to comprehend the world of daily life (*Lebenswelt*) from which these lives derive their significance, and the phenomenological approach



—its concern with essential structure, with founding relationships, with “sedimentation” of meaning, and with the uncovering of root presuppositions—provides the means by which social science may reconstruct mundane reality. At the same time, the use of *Verstehen* and of the subjective interpretation of meaning may lead to valid knowledge.

Although Schutz's work is profoundly indebted to Husserl and Max Weber, it has affinities with that of other European thinkers, Henri Bergson and Georg Simmel in particular. In the Anglo-American tradition, Schutz found the thought of Santayana, Whitehead, and James especially congenial. The sociologists who influenced him the most were Cooley, Thomas, and Mead. All these men stressed the activity of the actor and his action in relation to the experiential world. It was Schutz's distinctive achievement to develop these intellectual orientations by centering on the paradigm of action instead of the traditional theme of perception and by approaching both actor and action by way of the typifications of common-sense life. In the last decade, Schutz's work has aroused increasing interest among both philosophers and social scientists in the implications of phenomenology for a humanistic sociology and a theory of man.

MAURICE NATANSON

[See also KNOWLEDGE, SOCIOLOGY OF; SOCIAL STRUCTURE, article on SOCIAL STRUCTURAL ANALYSIS; VERSTEHEN; and the biographies of COOLEY; HUSSERL; MEAD; SIMMEL; THOMAS; WEBER, MAX.]

#### WORKS BY SCHUTZ

- (1932) 1967 *The Phenomenology of the Social World*. Evanston, Ill.: Northwestern University Press.  
*Collected Papers*. 3 vols. The Hague: Nijhoff, 1962–1965.  
 → Volume 1: *The Problem of Social Reality*. Volume 2: *Studies in Social Theory*. Volume 3: *Studies in Phenomenological Philosophy*.

#### SUPPLEMENTARY BIBLIOGRAPHY

- BERGER, PETER; and LUCKMANN, THOMAS 1966 *The Social Construction of Reality: A Treatise in the Sociology of Knowledge*. Garden City, N.Y.: Doubleday.  
 GURWITSCH, ARON (1957) 1965 *The Field of Consciousness*. Pittsburgh: Duquesne Univ. Press. → First published in French.  
 NATANSON, MAURICE (editor) 1963 *Philosophy of the Social Sciences: A Reader*. New York: Random House.  
 NATANSON, MAURICE 1966 *The Phenomenology of Alfred Schutz*. *Inquiry* 9:147–155.  
 SPIEGELBERG, HERBERT (1960) 1965 *The Phenomenological Movement: A Historical Introduction*. 2d ed. 2 vols. *Phaenomenologica*, Vols. 5–6. The Hague: Nijhoff.  
 STONIER, ALFRED; and BODE, KARL 1937 *A New Approach to the Methodology of the Social Sciences*. *Economica* New Series 4:406–424.  
 ZANER, RICHARD M. 1961 *Theory of Intersubjectivity*. *Alfred Schutz. Social Research* 28:71–93.

## SCIENCE

- I. THE HISTORY OF SCIENCE
- II. THE PHILOSOPHY OF SCIENCE
- III. THE SOCIOLOGY OF SCIENCE
- IV. SCIENCE—GOVERNMENT RELATIONS
- V. SCIENTISTS
- VI. SCIENTIFIC COMMUNICATION

Thomas S. Kuhn  
 Michael Scriven  
 Bernard Barber

Don K. Price  
 Warren O. Hagstrom  
 Norman Kaplan and  
 Norman W. Storer

### I

#### THE HISTORY OF SCIENCE

As an independent professional discipline, the history of science is a new field still emerging from a long and varied prehistory. Only since 1950, and initially only in the United States, has the majority of even its youngest practitioners been trained for, or committed to, a full-time scholarly career in the field. From their predecessors, most of whom were historians only by avocation and thus derived their goals and values principally from some other field, this younger generation inherits a constellation of sometimes irreconcilable objectives. The resulting tensions, though they have relaxed with increasing maturation of the profession, are still perceptible, particularly in the varied primary audiences to which the literature of the history of science continues to be addressed. Under the circumstances any brief report on development and current state is inevitably more personal and prognostic than for a longer-established profession.

**Development of the field.** Until very recently most of those who wrote the history of science were practicing scientists, sometimes eminent ones. Usually history was for them a by-product of pedagogy. They saw in it, besides intrinsic appeal, a means to elucidate the concepts of their specialty, to establish its tradition, and to attract students. The historical section with which so many technical treatises and monographs still open is contemporary illustration of what was for many centuries the primary form and exclusive source for the history of science. That traditional genre appeared in classical antiquity both in historical sections of technical treatises and in a few independent histories of the most developed ancient sciences, astronomy and mathematics. Similar works—together with a growing body of heroic biography—had a continuous history from the Renaissance through the eighteenth century, when their production was much stimulated by the Enlightenment's vision of science as at once the source and the exemplar of progress. From the last fifty years of that period come the earliest historical studies that are sometimes still used as such, among them the historical narratives embedded in

the technical works of Lagrange (mathematics) as well as the imposing separate treatises by Montucla (mathematics and physical science), Priestley (electricity and optics), and Delambre (astronomy). In the nineteenth and early twentieth centuries, though alternative approaches had begun to develop, scientists continued to produce both occasional biographies and magistral histories of their own specialties, for example, Kopp (chemistry), Poggendorff (physics), Sachs (botany), Zittel and Geikie (geology), and Klein (mathematics).

A second main historiographic tradition, occasionally indistinguishable from the first, was more explicitly philosophical in its objectives. Early in the seventeenth century Francis Bacon proclaimed the utility of histories of learning to those who would discover the nature and proper use of human reason. Condorcet and Comte are only the most famous of the philosophically inclined writers who, following Bacon's lead, attempted to base normative descriptions of true rationality on historical surveys of Western scientific thought. Before the nineteenth century this tradition remained predominantly programmatic, producing little significant historical research. But then, particularly in the writings of Whewell, Mach, and Duhem, philosophical concerns became a primary motive for creative activity in the history of science, and they have remained important since.

Both of these historiographic traditions, particularly when controlled by the textual-critical techniques of nineteenth-century German political history, produced occasional monuments of scholarship, which the contemporary historian ignores at his peril. But they simultaneously reinforced a concept of the field that has today been largely rejected by the nascent profession. The objective of these older histories of science was to clarify and deepen an understanding of *contemporary* scientific methods or concepts by displaying their evolution. Committed to such goals, the historian characteristically chose a single established science or branch of science—one whose status as sound knowledge could scarcely be doubted—and described when, where, and how the elements that in his day constituted its subject matter and presumptive method had come into being. Observations, laws, or theories which contemporary science had set aside as error or irrelevancy were seldom considered unless they pointed a methodological moral or explained a prolonged period of apparent sterility. Similar selective principles governed discussion of factors external to science. Religion, seen as a hindrance, and technology, seen as an occasional prerequisite to advance in instrumenta-

tion, were almost the only such factors which received attention. The outcome of this approach has recently been brilliantly parodied by the philosopher Joseph Agassi.

Until the early nineteenth century, of course, characteristics very much like these typified most historical writing. The romantics' passion for distant times and places had to combine with the scholarly standards of Biblical criticism before even general historians could be brought to recognize the interest and integrity of value systems other than their own. (The nineteenth century is, for example, the period when the Middle Ages were first observed to have a history.) That transformation of sensibility which most contemporary historians would suppose essential to their field was not, however, at once reflected in the history of science. Though they agreed about nothing else, both the romantic and the scientist-historian continued to view the development of science as a quasi-mechanical march of the intellect, the successive surrender of nature's secrets to sound methods skillfully deployed. Only in this century have historians of science gradually learned to see their subject matter as something different from a chronology of accumulating positive achievement in a technical specialty defined by hindsight. A number of factors contributed to this change.

Probably the most important was the influence, beginning in the late nineteenth century, of the history of philosophy. In that field only the most partisan could feel confident of his ability to distinguish positive knowledge from error and superstition. Dealing with ideas that had since lost their appeal, the historian could scarcely escape the force of an injunction which Bertrand Russell later phrased succinctly: "In studying a philosopher, the right attitude is neither reverence nor contempt, but first a kind of hypothetical sympathy, until it is possible to know what it feels like to believe in his theories." That attitude toward past thinkers came to the history of science from philosophy. Partly it was learned from men like Lange and Cassirer who dealt historically with people or ideas that were also important for scientific development. (Burt's *Metaphysical Foundations of Modern Physical Science* and Lovejoy's *Great Chain of Being* were, in this respect, especially influential.) And partly it was learned from a small group of Neo-Kantian epistemologists, particularly Brunschvicg and Meyerson, whose search for quasi-absolute categories of thought in older scientific ideas produced brilliant genetic analyses of concepts which the main tradition in the history of science had misunderstood or dismissed.

These lessons were reinforced by another deci-



sive event in the emergence of the contemporary profession. Almost a century after the Middle Ages had become important to the general historian, Pierre Duhem's search for the sources of modern science disclosed a tradition of medieval physical thought which, in contrast to Aristotle's physics, could not be denied an essential role in the transformation of physical theory that occurred in the seventeenth century. Too many of the elements of Galileo's physics and method were to be found there. But it was not possible, either, to assimilate it quite to Galileo's physics and to that of Newton, leaving the structure of the so-called Scientific Revolution unchanged but extending it greatly in time. The essential novelties of seventeenth-century science would be understood only if medieval science were explored first on its own terms and then as the base from which the "New Science" sprang. More than any other, that challenge has shaped the modern historiography of science. The writings which it has evoked since 1920, particularly those of E. J. Dijksterhuis, Anneliese Maier, and especially Alexandre Koyré, are the models which many contemporaries aim to emulate. In addition, the discovery of medieval science and its Renaissance role has disclosed an area in which the history of science can and must be integrated with more traditional types of history. That task has barely begun, but the pioneering synthesis by Butterfield and the special studies by Panofsky and Frances Yates mark a path which will surely be broadened and followed.

A third factor in the formation of the modern historiography of science has been a repeated insistence that the student of scientific development concern himself with positive knowledge as a whole and that general histories of science replace histories of special sciences. Traceable as a program to Bacon, and more particularly to Comte, that demand scarcely influenced scholarly performance before the beginning of this century, when it was forcefully reiterated by the universally venerated Paul Tannery and then put to practice in the monumental researches of George Sarton. Subsequent experience has suggested that the sciences are not, in fact, all of a piece and that even the superhuman erudition required for a general history of science could scarcely tailor their joint evolution to a coherent narrative. But the attempt has been crucial, for it has highlighted the impossibility of attributing to the past the divisions of knowledge embodied in contemporary science curricula. Today, as historians increasingly turn back to the detailed investigation of individual branches of science, they study fields which actually existed

in the periods that concern them, and they do so with an awareness of the state of other sciences at the time.

Still more recently, one other set of influences has begun to shape contemporary work in the history of science. Its result is an increased concern, deriving partly from general history and partly from German sociology and Marxist historiography, with the role of nonintellectual, particularly institutional and socioeconomic, factors in scientific development. Unlike the ones discussed above, however, these influences and the works responsive to them have to date scarcely been assimilated by the emerging profession. For all its novelties, the new historiography is still directed predominantly to the evolution of scientific ideas and of the tools (mathematical, observational, and experimental) through which these interact with each other and with nature. Its best practitioners have, like Koyré, usually minimized the importance of non-intellectual aspects of culture to the historical developments they consider. A few have acted as though the obtrusion of economic or institutional considerations into the history of science would be a denial of the integrity of science itself. As a result, there seem at times to be two distinct sorts of history of science, occasionally appearing between the same covers but rarely making firm or fruitful contact. The still dominant form, often called the "internal approach," is concerned with the substance of science as knowledge. Its newer rival, often called the "external approach," is concerned with the activity of scientists as a social group within a larger culture. Putting the two together is perhaps the greatest challenge now faced by the profession, and there are increasing signs of a response. Nevertheless, any survey of the field's present state must unfortunately still treat the two as virtually separate enterprises.

**Internal history.** What are the maxims of the new internal historiography? Insofar as possible (it is never entirely so, nor could history be written if it were), the historian should set aside the science that he knows. His science should be learned from the textbooks and journals of the period he studies, and he should master these and the indigenous traditions they display before grappling with innovators whose discoveries or inventions changed the direction of scientific advance. Dealing with innovators, the historian should try to think as they did. Recognizing that scientists are often famous for results they did not intend, he should ask what problems his subject worked at and how these became problems for him. Recognizing that a historic discovery is rarely

quite the one attributed to its author in later textbooks (pedagogic goals inevitably transform a narrative), the historian should ask what his subject thought he had discovered and what he took the basis of that discovery to be. And in this process of reconstruction the historian should pay particular attention to his subject's apparent errors, not for their own sake but because they reveal far more of the mind at work than do the passages in which a scientist seems to record a result or an argument that modern science still retains.

For at least thirty years the attitudes which these maxims are designed to display have increasingly guided the best interpretive scholarship in the history of science, and it is with scholarship of that sort that this article is predominantly concerned. (There are other types, of course, though the distinction is not sharp, and much of the most worthwhile effort of historians of science is devoted to them. But this is not the place to consider work like that of, say, Needham, Neugebauer, and Thorndike, whose indispensable contribution has been to establish and make accessible texts and traditions previously known only through myth.) Nevertheless, the subject matter is immense; there have been few professional historians of science (in 1950 scarcely more than half a dozen in the United States); and their choice of topic has been far from random. There remain vast areas for which not even the basic developmental lines are clear.

Probably because of their special prestige, physics, chemistry, and astronomy dominate the historical literature of science. But even in these fields effort has been unevenly distributed, particularly in this century. Because they sought contemporary knowledge in the past, the nineteenth-century scientist-historians compiled surveys which often ranged from antiquity to their own day or close to it. In the twentieth century a few scientists, like Dugas, Jammer, Partington, Truesdell, and Whitaker, have written from a similar viewpoint, and some of their surveys carry the history of special fields close to the present. But few practitioners of the most developed sciences still write histories, and the members of the emerging profession have up to this time been far more systematically and narrowly selective, with a number of unfortunate consequences. The deep and sympathetic immersion in the sources which their work demands virtually prohibits wide-ranging surveys, at least until more of the field has been examined in depth. Starting with a clean slate, as they at least feel they are, this group naturally tries first to establish the early phases of a science's development,

and few get beyond that point. Besides, until the last few years almost no member of the new group has had sufficient command of the science (particularly mathematics, usually the decisive hurdle) to become a vicarious participant in the more recent research of the technically most developed disciplines.

As a result, though the situation is now changing rapidly with the entry both of more and of better-prepared people into the field, the recent literature of the history of science tends to end at the point where the technical source materials cease to be accessible to a man with elementary college scientific training. There are fine studies of mathematics to Leibniz (Boyer, Michel); of astronomy and mechanics to Newton (Clagett, Costabel, Dijksterhuis, Koyré, and Maier), of electricity to Coulomb (Cohen), and of chemistry to Dalton (Boas, Crosland, Daumas, Guerlac, Metzger). But almost no work within the new tradition has as yet been published on the mathematical physical science of the eighteenth century or on any physical science in the nineteenth.

For the biological and earth sciences, the literature is even less well developed, partly because only those subspecialties which, like physiology, relate closely to medicine had achieved professional status before the late nineteenth century. There are few of the older surveys by scientists, and the members of the new profession are only now beginning in any number to explore these fields. In biology at least there is prospect of rapid change, but up to this point the only areas much studied are nineteenth-century Darwinism and the anatomy and physiology of the sixteenth and seventeenth centuries. On the second of these topics, however, the best of the book-length studies (e.g., O'Malley and Singer) deal usually with special problems and persons and thus scarcely display an evolving scientific tradition. The literature on evolution, in the absence of adequate histories of the technical specialties which provided Darwin with both data and problems, is written at a level of philosophical generality which makes it hard to see how his *Origin of Species* could have been a major achievement, much less an achievement in the sciences. Dupree's model study of the botanist Asa Gray is among the few noteworthy exceptions.

As yet the new historiography has not touched the social sciences. In these fields the historical literature, where it exists, has been produced entirely by practitioners of the science concerned, Boring's *History of Experimental Psychology* being perhaps the outstanding example. Like the older histories of the physical sciences, this literature is



often indispensable, but as history it shares their limitations. (The situation is typical for relatively new sciences: practitioners in these fields are ordinarily expected to know about the development of their specialties, which thus regularly acquire a quasi-official history; thereafter something very like Gresham's law applies.) This area therefore offers particular opportunities both to the historian of science and, even more, to the general intellectual or social historian, whose background is often especially appropriate to the demands of these fields. The preliminary publications of Stocking on the history of American anthropology provide a particularly fruitful example of the perspective which the general historian can apply to a scientific field whose concepts and vocabulary have only very recently become esoteric.

**External history.** Attempts to set science in a cultural context which might enhance understanding both of its development and of its effects have taken three characteristic forms, of which the oldest is the study of scientific institutions. Bishop Sprat prepared his pioneering history of the Royal Society of London almost before that organization had received its first charter, and there have since been innumerable in-house histories of individual scientific societies. These books are, however, useful principally as source materials for the historian, and only in this century have students of scientific development started to make use of them. Simultaneously they have begun seriously to examine the other types of institutions, particularly educational, which may promote or inhibit scientific advance. As elsewhere in the history of science, most of the literature on institutions deals with the seventeenth century. The best of it is scattered through periodicals (the once standard book-length accounts are regrettably out-of-date) from which it can be retrieved, together with much else concerning the history of science, through the annual "Critical Bibliography" of the journal *Isis* and through the quarterly *Bulletin signalétique* of the Centre National de la Recherche Scientifique, Paris. Guerlac's classic study on the professionalization of French chemistry, Schofield's history of the Lunar Society, and a recent collaborative volume (Taton) on scientific education in France are among the very few works on eighteenth-century scientific institutions. For the nineteenth, only Cardwell's study of England, Dupree's of the United States, and Vucinich's of Russia begin to replace the fragmentary but immensely suggestive remarks scattered, often in footnotes, through the first volume of Merz's *History of European Thought in the Nineteenth Century*.

Intellectual historians have frequently considered the impact of science on various aspects of Western thought, particularly during the seventeenth and eighteenth centuries. For the period since 1700, however, these studies are peculiarly unsatisfying insofar as they aim to demonstrate the influence, and not merely the prestige, of science. The name of a Bacon, a Newton, or a Darwin is a potent symbol: there are many reasons to invoke it besides recording a substantive debt. And the recognition of isolated conceptual parallels, e.g., between the forces that keep a planet in its orbit and the system of checks and balances in the U.S. constitution, more often demonstrates interpretive ingenuity than the influence of science on other areas of life. No doubt scientific concepts, particularly those of broad scope, do help to change extrascientific ideas. But the analysis of their role in producing this kind of change demands immersion in the literature of science. The older historiography of science does not, by its nature, supply what is needed, and the new historiography is too recent and its products too fragmentary to have had much effect. Though the gap seems small, there is no chasm that more needs bridging than that between the historian of ideas and the historian of science. Fortunately there are a few works to point the way. Among the more recent are Nicolson's pioneering studies of science in seventeenth- and eighteenth-century literature, Westfall's discussion of natural religion, Gillispie's chapter on science in the Enlightenment, and Roger's monumental survey of the role of the life sciences in eighteenth-century French thought.

The concern with institutions and that with ideas merge naturally in a third approach to scientific development. This is the study of science in a geographical area too small to permit concentration on the evolution of any particular technical specialty but sufficiently homogeneous to enhance an understanding of science's social role and setting. Of all the types of external history, this is the newest and most revealing, for it calls forth the widest range of historical and sociological experience and skill. The small but rapidly growing literature on science in America (Dupree, Hindle, Shryock) is a prominent example of this approach, and there is promise that current studies of science in the French Revolution may yield similar illumination. Merz, Lilley, and Ben-David point to aspects of the nineteenth century on which much similar effort must be expended. The topic which has, however, evoked the greatest activity and attention is the development of science in seventeenth-century England. Because it has become the

center of vociferous debate both about the origin of modern science and about the nature of the history of science, this literature is an appropriate focus for separate discussion. Here it stands for a type of research: the problems it presents will provide perspective on the relations between the internal and external approaches to the history of science.

**The Merton thesis.** The most visible issue in the debate about seventeenth-century science has been the so-called Merton thesis, really two overlapping theses with distinguishable sources. Both aim ultimately to account for the special productivity of seventeenth-century science by correlating its novel goals and values—summarized in the program of Bacon and his followers—with other aspects of contemporary society. The first, which owes something to Marxist historiography, emphasizes the extent to which the Baconians hoped to learn from the practical arts and in turn to make science useful. Repeatedly they studied the techniques of contemporary craftsmen—glassmakers, metallurgists, mariners, and the like—and many also devoted at least a portion of their attention to pressing practical problems of the day, e.g., those of navigation, land drainage, and deforestation. The new problems, data, and methods fostered by these novel concerns are, Merton supposes, a principal reason for the substantive transformation experienced by a number of sciences during the seventeenth century. The second thesis points to the same novelties of the period but looks to Puritanism as their primary stimulant. (There need be no conflict. Max Weber, whose pioneering suggestion Merton was investigating, had argued that Puritanism helped to legitimize a concern with technology and the useful arts.) The values of settled Puritan communities—for example, an emphasis upon justification through works and on direct communion with God through nature—are said to have fostered both the concern with science and the empirical, instrumental, and utilitarian tone which characterized it during the seventeenth century.

Both of these theses have since been extended and also attacked with vehemence, but no consensus has emerged. (An important confrontation, centering on papers by Hall and de Santillana, appears in the symposium of the Institute for the History of Science edited by Clagett; Zilsel's pioneering paper on William Gilbert can be found in the collection of relevant articles from the *Journal of the History of Ideas* edited by Wiener and Noland. Most of the rest of the literature, which is voluminous, can be traced through the

footnotes in a recently published controversy over the work of Christopher Hill.) In this literature the most persistent criticisms are those directed to Merton's definition and application of the label "Puritan," and it now seems clear that no term so narrowly doctrinal in its implications will serve. Difficulties of this sort can surely be eliminated, however, for the Baconian ideology was neither restricted to scientists nor uniformly spread through all classes and areas of Europe. Merton's label may be inadequate, but there is no doubt that the phenomenon he describes did exist. The more significant arguments against his position are the residual ones which derive from the recent transformation in the history of science. Merton's image of the Scientific Revolution, though long-standing, was rapidly being discredited as he wrote, particularly in the role it attributed to the Baconian movement.

Participants in the older historiographic tradition did sometimes declare that science as they conceived it owed nothing to economic values or religious doctrine. Nevertheless, Merton's emphases on the importance of manual work, experimentation, and the direct confrontation with nature were familiar and congenial to them. The new generation of historians, in contrast, claims to have shown that the radical sixteenth- and seventeenth-century revisions of astronomy, mathematics, mechanics, and even optics owed very little to new instruments, experiments, or observations. Galileo's primary method, they argue, was the traditional thought experiment of scholastic science brought to a new perfection. Bacon's naive and ambitious program was an impotent delusion from the start. The attempts to be useful failed consistently; the mountains of data provided by new instruments were of little assistance in the transformation of existing scientific theory. If cultural novelties are required to explain why men like Galileo, Descartes, and Newton were suddenly able to see well-known phenomena in a new way, those novelties are predominantly intellectual and include Renaissance Neoplatonism, the revival of ancient atomism, and the rediscovery of Archimedes. Such intellectual currents were, however, at least as prevalent and productive in Roman Catholic Italy and France as in Puritan circles in Britain or Holland. And nowhere in Europe, where these currents were stronger among courtiers than among craftsmen, do they display a significant debt to technology. If Merton were right, the new image of the Scientific Revolution would apparently be wrong.

In their more detailed and careful versions,



which include essential qualification, these arguments are entirely convincing, up to a point. The men who transformed scientific theory during the seventeenth century sometimes talked like Baconians, but it has yet to be shown that the ideology which a number of them embraced had a major effect, substantive or methodological, on their central contributions to science. Those contributions are best understood as the result of the internal evolution of a cluster of fields which, during the sixteenth and seventeenth centuries, were pursued with renewed vigor and in a new intellectual milieu. That point, however, can be relevant only to the revision of the Merton thesis, not to its rejection. One aspect of the ferment which historians have regularly labeled "the Scientific Revolution" was a radical programmatic movement centering in England and the Low Countries, though it was also visible for a time in Italy and France. That movement, which even the present form of Merton's argument does make more comprehensible, drastically altered the appeal, the locus, and the nature of much scientific research during the seventeenth century, and the changes have been permanent. Very likely, as contemporary historians argue, none of these novel features played a large role in transforming scientific concepts during the seventeenth century, but historians must learn to deal with them nonetheless. Perhaps the following suggestions, whose more general import will be considered in the next section, may prove helpful.

Omitting the biological sciences, for which close ties to medical crafts and institutions dictate a more complex developmental pattern, the main branches of science transformed during the sixteenth and seventeenth centuries were astronomy, mathematics, mechanics, and optics. It is their development which makes the Scientific Revolution seem a revolution in concepts. Significantly, however, this cluster of fields consists exclusively of classical sciences. Highly developed in antiquity, they found a place in the medieval university curriculum where several of them were significantly further developed. Their seventeenth-century metamorphosis, in which university-based men continued to play a significant role, can reasonably be portrayed as primarily an extension of an ancient and medieval tradition developing in a new conceptual environment. Only occasionally need one have recourse to the Baconian programmatic movement when explaining the transformation of these fields.

By the seventeenth century, however, these were not the only areas of intense scientific activity, and the others—among them the study of electricity

and magnetism, of chemistry, and of thermal phenomena—display a different pattern. As sciences, as fields to be scrutinized systematically for an increased understanding of nature, they were all novelties during the Scientific Revolution. Their main roots were not in the learned university tradition but often in the established crafts, and they were all critically dependent both on the new program of experimentation and on the new instrumentation which craftsmen often helped to introduce. Except occasionally in medical schools, they rarely found a place in universities before the nineteenth century, and they were meanwhile pursued by amateurs loosely clustered around the new scientific societies that were the institutional manifestation of the Scientific Revolution. Obviously these are the fields, together with the new mode of practice they represent, which a revised Merton thesis may help us understand. Unlike that in the classical sciences, research in these fields added little to man's understanding of nature during the seventeenth century, a fact which has made them easy to ignore when evaluating Merton's viewpoint. But the achievements of the late eighteenth and of the nineteenth centuries will not be comprehensible until they are taken fully into account. The Baconian program, if initially barren of conceptual fruits, nevertheless inaugurated a number of the major modern sciences.

**Internal and external history.** Because they underscore distinctions between the earlier and later stages of a science's evolution, these remarks about the Merton thesis illustrate aspects of scientific development recently discussed in a more general way by Kuhn. Early in the development of a new field, he suggests, social needs and values are a major determinant of the problems on which its practitioners concentrate. Also during this period, the concepts they deploy in solving problems are extensively conditioned by contemporary common sense, by a prevailing philosophical tradition, or by the most prestigious contemporary sciences. The new fields which emerged in the seventeenth century and a number of the modern social sciences provide examples. Kuhn argues, however, that the later evolution of a technical specialty is significantly different in ways at least foreshadowed by the development of the classical sciences during the Scientific Revolution. The practitioners of a mature science are men trained in a sophisticated body of traditional theory and of instrumental, mathematical, and verbal technique. As a result, they constitute a special subculture, one whose members are the exclusive audience for, and judges of, each other's work. The problems on which such

specialists work are no longer presented by the external society but by an internal challenge to increase the scope and precision of the fit between existing theory and nature. And the concepts used to resolve these problems are normally close relatives of those supplied by prior training for the specialty. In short, compared with other professional and creative pursuits, the practitioners of a mature science are effectively insulated from the cultural milieu in which they live their extra-professional lives.

That quite special, though still incomplete, insulation is the presumptive reason why the internal approach to the history of science, conceived as autonomous and self-contained, has seemed so nearly successful. To an extent unparalleled in other fields, the development of an individual technical specialty can be understood without going beyond the literature of that specialty and a few of its near neighbors. Only occasionally need the historian take note of a particular concept, problem, or technique which entered the field from outside. Nevertheless, the apparent autonomy of the internal approach is misleading in essentials, and the passion sometimes expended in its defense has obscured important problems. The insulation of a mature scientific community suggested by Kuhn's analysis is an insulation primarily with respect to concepts and secondarily with respect to problem structure. There are, however, other aspects of scientific advance, such as its timing. These do depend critically on the factors emphasized by the external approach to scientific development. Particularly when the sciences are viewed as an interacting group rather than as a collection of specialties, the cumulative effects of external factors can be decisive.

Both the attraction of science as a career and the differential appeal of different fields are, for example, significantly conditioned by factors external to science. Furthermore, since progress in one field is sometimes dependent on the prior development of another, differential growth rates may affect an entire evolutionary pattern. Similar considerations, as noted above, play a major role in the inauguration and initial form of new sciences. In addition, a new technology or some other change in the conditions of society may selectively alter the felt importance of a specialty's problems or even create new ones for it. By doing so they may sometimes accelerate the discovery of areas in which an established theory ought to work but does not, thereby hastening its rejection and replacement by a new one. Occasionally, they may even shape the substance of that new theory

by ensuring that the crisis to which it responds occurs in one problem area rather than another. Or again, through the crucial intermediary of institutional reform, external conditions may create new channels of communication between previously disparate specialties, thus fostering cross-fertilization which would otherwise have been absent or long delayed.

There are numerous other ways, including direct subsidy, in which the larger culture impinges on scientific development, but the preceding sketch should sufficiently display a direction in which the history of science must now develop. Though the internal and external approaches to the history of science have a sort of natural autonomy, they are, in fact, complementary concerns. Until they are practiced as such, each drawing from the other, important aspects of scientific development are unlikely to be understood. That mode of practice has hardly yet begun, as the response to the Merton thesis indicates, but perhaps the analytic categories it demands are becoming clear.

The relevance of the history of science. Turning in conclusion to the question about which judgments must be the most personal of all, one may ask about the potential harvest to be reaped from the work of this new profession. First and foremost will be more and better histories of science. Like any other scholarly discipline, the field's primary responsibility must be to itself. Increasing signs of its selective impact on other enterprises may, however, justify brief analysis.

Among the areas to which the history of science relates, the one least likely to be significantly affected is scientific research itself. Advocates of the history of science have occasionally described their field as a rich repository of forgotten ideas and methods, a few of which might well dissolve contemporary scientific dilemmas. When a new concept or theory is successfully deployed in a science, some previously ignored precedent is usually discovered in the earlier literature of the field. It is natural to wonder whether attention to history might not have accelerated the innovation. Almost certainly, however, the answer is no. The quantity of material to be searched, the absence of appropriate indexing categories, and the subtle but usually vast differences between the anticipation and the effective innovation, all combine to suggest that reinvention rather than rediscovery will remain the most efficient source of scientific novelty.

The more likely effects of the history of science on the fields it chronicles are indirect, providing increased understanding of the scientific enterprise itself. Though a clearer grasp of the nature of



scientific development is unlikely to resolve particular puzzles of research, it may well stimulate reconsideration of such matters as science education, administration, and policy. Probably, however, the implicit insights which historical study can produce will first need to be made explicit by the intervention of other disciplines, of which three now seem particularly likely to be effective.

Though the intrusion still evokes more heat than light, the philosophy of science is today the field in which the impact of the history of science is most apparent. Feyerabend, Hanson, Hesse, and Kuhn have all recently insisted on the inappropriateness of the traditional philosopher's ideal image of science, and in search of an alternative they have all drawn heavily from history. Following directions pointed by the classic statements of Norman Campbell and Karl Popper (and sometimes also significantly influenced by Ludwig Wittgenstein), they have at least raised problems that the philosophy of science is no longer likely to ignore. The resolution of those problems is for the future, perhaps for the indefinitely distant future. There is as yet no developed and matured "new philosophy" of science. But already the questioning of older stereotypes, mostly positivistic, is proving a stimulus and release to some practitioners of those newer sciences which have most depended upon explicit canons of scientific method in their search for professional identity.

A second field in which the history of science is likely to have increasing effect is the sociology of science. Ultimately neither the concerns nor the techniques of that field need be historical. But in the present underdeveloped state of their specialty, sociologists of science can well learn from history something about the shape of the enterprise they investigate. The recent writings of Ben-David, Hagstrom, Merton, and others give evidence that they are doing so. Very likely it will be through sociology that the history of science has its primary impact on science policy and administration.

Closely related to the sociology of science (perhaps equivalent to it if the two are properly construed) is a field that, though it scarcely yet exists, is widely described as "the science of science." Its goal, in the words of its leading exponent, Derek Price, is nothing less than "the theoretic analysis of the structure and behavior of science itself," and its techniques are an eclectic combination of the historian's, the sociologist's, and the econometrician's. No one can yet guess to what extent that goal is attainable, but any progress toward it will inevitably and immediately enhance the signifi-

cance both to social scientists and to society of continuing scholarship in the history of science.

THOMAS S. KUHN

[Other relevant material may be found in the biographies of KOYRÉ and SARTON.]

#### BIBLIOGRAPHY

- AGASSI, JOSEPH 1963 *Towards an Historiography of Science. History and Theory*, Beiheft 2. The Hague: Mouton.
- BEN-DAVID, JOSEPH 1960 Scientific Productivity and Academic Organization in Nineteenth-century Medicine. *American Sociological Review* 25:828-843.
- BOAS, MARIE 1958 Robert Boyle and Seventeenth-century Chemistry. Cambridge Univ. Press.
- BOYER, CARL B. (1939) 1949 *The Concepts of the Calculus: A Critical and Historical Discussion of the Derivative and the Integral*. New York: Hafner. → A paperback edition was published in 1959 by Dover as *The History of the Calculus and Its Conceptual Development*.
- BUTTERFIELD, HERBERT (1950) 1957 *The Origins of Modern Science, 1300-1800*. 2d ed., rev. New York: Macmillan. → A paperback edition was published in 1962 by Collier.
- CARDWELL, DONALD S. L. 1957 *The Organisation of Science in England: A Retrospect*. Melbourne and London: Heinemann.
- CLAGETT, MARSHALL 1959 *The Science of Mechanics in the Middle Ages*. Madison: Univ. of Wisconsin Press.
- COHEN, I. BERNARD 1956 *Franklin and Newton: An Inquiry Into Speculative Newtonian Experimental Science and Franklin's Work in Electricity as an Example Thereof*. American Philosophical Society, Memoirs, Vol. 43. Philadelphia: The Society.
- COSTABEL, PIERRE 1960 *Leibniz et la dynamique: Les textes de 1692*. Paris: Hermann.
- CROSLAND, MAURICE 1963 *The Development of Chemistry in the Eighteenth Century. Studies on Voltaire and the Eighteenth Century* 24:369-441.
- DAUMAS, MAURICE 1955 *Lavoisier: Théoricien et expérimentateur*. Paris: Presses Universitaires de France.
- DIJKSTERHUIS, EDWARD J. (1950) 1961 *The Mechanization of the World Picture*. Oxford: Clarendon. → First published in Dutch.
- DUGAS, RENÉ (1950) 1955 *A History of Mechanics*. Neuchâtel (Switzerland): Éditions du Griffon; New York: Central Book. → First published in French.
- DUEHM, PIERRE 1906-1913 *Études sur Léonard de Vinci*. 3 vols. Paris: Hermann.
- DUPREE, A. HUNTER 1957 *Science in the Federal Government: A History of Policies and Activities to 1940*. Cambridge, Mass.: Belknap.
- DUPREE, A. HUNTER 1959 *Asa Gray: 1810-1888*. Cambridge, Mass.: Harvard Univ. Press.
- FEYERABEND, P. K. 1962 Explanation, Reduction and Empiricism. Pages 28-97 in Herbert Feigl and Grover Maxwell (editors), *Scientific Explanation, Space, and Time*. Minnesota Studies in the Philosophy of Science, Vol. 3. Minneapolis: Univ. of Minnesota Press.
- GILLISPIE, CHARLES C. 1960 *The Edge of Objectivity: An Essay in the History of Scientific Ideas*. Princeton Univ. Press.

- GUERLAC, HENRY 1959 Some French Antecedents of the Chemical Revolution. *Chymia* 5:73-112.
- GUERLAC, HENRY 1961 *Lavoisier; the Crucial Year: The Background and Origin of His First Experiments on Combustion in 1772*. Ithaca, N.Y.: Cornell Univ. Press.
- HAGSTROM, WARREN O. 1965 *The Scientific Community*. New York: Basic Books.
- HANSON, NORWOOD R. (1958) 1961 *Patterns of Discovery: An Inquiry Into the Conceptual Foundations of Science*. Cambridge Univ. Press.
- HESSE, MARY B. 1963 *Models and Analogies in Science*. London: Sheed & Ward.
- HILL, CHRISTOPHER 1965 Debate: Puritanism, Capitalism and the Scientific Revolution. *Past and Present* No. 29:88-97. → Articles relevant to the debate may also be found in numbers 28, 31, 32, and 33.
- HINDLE, BROOKE 1956 *The Pursuit of Science in Revolutionary America, 1735-1789*. Chapel Hill: Univ. of North Carolina Press.
- INSTITUTE FOR THE HISTORY OF SCIENCE, UNIVERSITY OF WISCONSIN, 1957 1959 *Critical Problems in the History of Science: Proceedings*. Edited by Marshall Clagett. Madison: Univ. of Wisconsin Press.
- JAMMER, MAX 1961 *Concepts of Mass in Classical and Modern Physics*. Cambridge, Mass.: Harvard Univ. Press.
- JOURNAL OF THE HISTORY OF IDEAS 1957 *Roots of Scientific Thought: A Cultural Perspective*. Edited by Philip P. Wiener and Aaron Noland. New York: Basic Books. → Selections from the first 18 volumes of the *Journal*.
- KOYRÉ, ALEXANDRE 1939 *Études galiléennes*. 3 vols. Actualités scientifiques et industrielles, Nos. 852, 853, and 854. Paris: Hermann. → Volume 1: *À l'aube de la science classique*. Volume 2: *La loi de la chute des corps: Descartes et Galilée*. Volume 3: *Galilée et la loi d'inertie*.
- KOYRÉ, ALEXANDRE 1961 *La révolution astronomique: Copernic, Kepler, Borelli*. Paris: Hermann.
- KUHN, THOMAS S. 1962 *The Structure of Scientific Revolutions*. Univ. of Chicago Press. → A paperback edition was published in 1964.
- LILLEY, S. 1949 Social Aspects of the History of Science. *Archives internationales d'histoire des sciences* 2:376-443.
- MAIER, ANNELIESE 1949-1958 *Studien zur Naturphilosophie der Spätscholastik*. 5 vols. Rome: Edizioni di "Storia e Letteratura."
- MERTON, ROBERT K. (1938) 1967 *Science, Technology and Society in Seventeenth-century England*. New York: Fertig.
- MERTON, ROBERT K. 1957 Priorities in Scientific Discovery: A Chapter in the Sociology of Science. *American Sociological Review* 22:635-659.
- METZGER, HÉLÈNE 1930 *Newton, Stahl, Boerhaave et la doctrine chimique*. Paris: Alcan.
- MEYERSON, ÉMILE (1908) 1964 *Identity and Reality*. London: Allen & Unwin. → First published in French.
- MICHEL, PAUL-HENRI 1950 *De Pythagore à Euclide*. Paris: Édition "Les Belles Lettres."
- NEEDHAM, JOSEPH 1954-1965 *Science and Civilization in China*. 4 vols. Cambridge Univ. Press.
- NEUGEBAUER, OTTO (1951) 1957 *The Exact Sciences in Antiquity*. 2d ed. Providence, R.I.: Brown Univ. Press. → A paperback edition was published in 1962 by Harper.
- NICOLSON, MARJORIE H. (1950) 1960 *The Breaking of the Circle: Studies in the Effect of the "New Science" Upon Seventeenth-century Poetry*. Rev. ed. New York: Columbia Univ. Press. → A paperback edition was published in 1962.
- O'MALLEY, CHARLES D. 1964 *Andreas Vesalius of Brussels, 1514-1564*. Berkeley and Los Angeles: Univ. of California Press.
- PANOFSKY, ERWIN 1954 *Galileo as a Critic of the Arts*. The Hague: Nijhoff.
- PARTINGTON, JAMES R. 1962— *A History of Chemistry*. New York: St. Martins. → Volumes 2-4 were published from 1962 to 1964; Volume 1 is in preparation.
- PRICE, DEREK J. DE Solla 1966 *The Science of Scientists. Medical Opinion and Review* 1:81-97.
- ROGER, JACQUES 1963 *Les sciences de la vie dans la pensée française du XVIII<sup>e</sup> siècle: La génération des animaux de Descartes à l'Encyclopédie*. Paris: Colin.
- SARTON, GEORGE 1927-1948 *Introduction to the History of Science*. 3 vols. Baltimore: Williams & Wilkins.
- SCHOFIELD, ROBERT E. 1963 *The Lunar Society of Birmingham: A Social History of Provincial Science and Industry in Eighteenth-century England*. Oxford: Clarendon.
- SHRYOCK, RICHARD H. (1936) 1947 *The Development of Modern Medicine*. 2d ed. New York: Knopf.
- SINGER, CHARLES J. 1922 *The Discovery of the Circulation of the Blood*. London: Bell.
- STOCKING, GEORGE W. JR. 1966 Franz Boas and the Culture Concept in Historical Perspective. *American Anthropologist New Series* 68:867-882.
- TATON, RENÉ (editor) 1964 *Enseignement et diffusion des sciences en France au XVIII<sup>e</sup> siècle*. Paris: Hermann.
- THORNDIKE, LYNN (1923-1958) 1959-1964 *A History of Magic and Experimental Science*. 8 vols. New York: Columbia Univ. Press.
- TRUESDELL, CLIFFORD A. 1960 *The Rational Mechanics of Flexible or Elastic Bodies 1638-1788: Introduction to Leonhardi Euleri Opera omnia Vol. X et XI seriei secundae. Leonhardi Euleri Opera omnia, Ser. 2, Vol. 11, part 2. Turin (Italy): Fussli*.
- VUCINICH, ALEXANDER S. 1963 *Science in Russian Culture*. Volume 1: *A History to 1860*. Stanford Univ. Press.
- WESTFALL, RICHARD S. 1958 *Science and Religion in Seventeenth-century England*. New Haven: Yale Univ. Press.
- WHITTAKER, EDMUND 1951-1953 *A History of the Theories of Aether and Electricity*. 2 vols. London: Nelson. → Volume 1: *The Classical Theories*. Volume 2: *The Modern Theories, 1900-1926*. Volume 1 is a revised edition of *A History of the Theories of Aether and Electricity From the Age of Descartes to the Close of the Nineteenth Century*, published in 1910. A paperback edition was published in 1960 by Harper.
- YATES, FRANCES A. 1964 *Giordano Bruno and the Hermetic Tradition*. Univ. of Chicago Press.

## II

## THE PHILOSOPHY OF SCIENCE

The general topic of the philosophy of science can be divided into subareas by subject matter: the philosophy of physics, the philosophy of biology, the philosophy of the social sciences, and so



on. But it may also be divided into discussion of *structural* problems in science, on the one hand, and *substantive* problems within specified sciences, on the other. Structural problems are those traditionally associated with such topics as scientific inference, classification, explanation, prediction, measurement, probability, and determinism. Substantive problems arise when a question is asked such as "How can learning theory be axiomatized?" or "How should anxiety be defined?" Substantive problems, as a class, merge gradually into abstract scientific problems, which are properly handled by scientists themselves.

The substantive philosophy of science usually requires some professional training in the relevant field, whereas the structural philosophy of science may require only a general scientific education. Both require professional training in philosophy or, to be more specific, in logic, broadly conceived (that is, not only in classical or in symbolic logic). The history of science is also of very great importance to the philosopher of science, for it provides him with more numerous and more diverse examples of conceptual schemes and theories than he could ever find in contemporary science, and such schemes and theories are his raw data.

It will be seen, then, that a large part of the philosophy of science stands to scientific theorizing, classifying, and so on, as experimental design stands to scientific experimenting. For the philosophy of science is simply the discussion of general criteria for theories, classifications, and the rest, and is necessary as long as the intuitions or assertions of scientists about what constitutes good procedure are in conflict or might be improved by analysis. Curiously enough, however, many scientists reject the philosophy of science as irrelevant to their own activities, although they constantly talk it and teach it and illustrate its relevance in their work, sometimes under the title "methodology" and sometimes just as advice without a title. It is only the hyperspecialized, rock-ribbed empiricist who can entirely avoid discussing the subject in his scientific work and teaching, and even he usually contradicts himself by espousing operationism as the epitome of empiricism—a rather elementary mistake which would have grave consequences for the validity and utility of his results if he acted on it. Often the antagonism of science to philosophy seems due to a low tolerance of ambiguity, and particularly to insecurity about involvement with the more debatable and less decidable issues of philosophy. This is surely an immature reaction, unless it can be shown that

these issues need not be raised in the teaching and practice of science.

This article will provide (a) a brief discussion of those aspects of the structural philosophy of science which have special relevance for the social sciences and (b) some reference to substantive issues in the philosophy of particular social sciences. In general, these aspects will be intertwined, in order to illustrate the immediate relevance of the structural topics to substantive issues.

**Main concerns of the philosophy of science.** Science has been said to be concerned with observation, description, definition, classification, measurement, experimentation, generalization, explanation, prediction, evaluation, and control of the world. This list is of course much too comprehensive; to be at all useful it has to be narrowed down in the course of examining individual scientific activities (for instance, instead of analyzing description in general, we might ask what *kind* of describing is scientific, as opposed to poetic). In any case, there are as many (though probably no more) points of view about main topics in the philosophy of science as about main topics in any of the sciences. If the area of dispute sometimes appears greater than it is, this is because when a subtopic in philosophy is settled, it is given a new name and called a science. Philosophy is the mother of sciences and has spawned physics, astronomy, symbolic logic, economics, and psychology, among other subjects. The domain of fundamental and still unresolved conceptual disputes at any time is called philosophy and, of course, bears the cross of conflict. But, perhaps for that very reason, it is a residue that eternally regenerates, is perpetually fertile. The social sciences began with a similar handicap; what was well known about human behavior was called common sense, and only the muddy residue was left for psychology and the other social sciences to clarify. It is no wonder that a subject is difficult and debatable when it is defined to exclude the easy and certain.

Because "science" is a term for an activity as well as for a body of knowledge, we ought not to be surprised if it is continuous with the prescientific activities directed to the same ends, and if these, in turn, are an extension of prelinguistic adaptive behavior. This synoptic view of man's cognitive development and current equipment should—and, I believe, does—provide some criteria for good scientific practice. We can begin by applying it to the various procedures connected with scientific concept formation: observation, description, definition, and classification, with meas-

urement and generalization as closely related activities.

**Concept clarification.** It is now widely accepted in the philosophy of science that we must qualify or reject the traditional idea that the basic data from which the scientist forms his hypotheses and classifications are provided by a faculty of observation which is somehow independent of science. Not only has the attempt by sense-data theorists and phenomenologists to identify these basic data been a total failure, but the failure begins to appear explicable. We have come to see the "observation language"—the nonscientific language in which such basic data were supposed to be reported—as really just a "theory language," with somewhat higher interjudge reliability than more abstract language, yet still subject to perceptual errors in its application and even to pervasive errors caused by infiltration by unnoticed theory. When errors of the latter kind are recognized, we may redescribe our observations or we may retain the old language and simply reinterpret it, perhaps metaphysically. Thus Feyerabend has pointed out (in conversation with the author) that the expression "The sun is rising," although it originally incorporated a false astronomical theory, has been retained as a statement of pure observation. So there is no absolutely basic observation level on which all theory rests and, even more clearly, no sharp distinction between observation and theories. We must think instead of a constant interaction and exchange between observation, which seems immune from correction, and speculative theory. Indeed, the development of new instruments may simply render observable the hitherto unobservable (as when the electron microscope was developed) or the observed mistaken (as when it became possible to measure the curvature of the seemingly flat earth). Does this relativization of the distinction between observation and theory destroy the distinction's utility? Not at all. Indeed, it strengthens our understanding of the hierarchical structure of science as a whole when we realize that observables of the  $n$ th level (for instance, in neurophysiology) may include some hypothetical constructs from the  $(n + 1)$ th level (in this case, individual psychology).

What leads us, then, to introduce concepts at a given level? Not just the pressure of their existence on our blank brain. Rather, the preconscious selective effect of certain environmental and internal pressures began the process of concept formation in our ancestors, as in other organisms—a sequence of events that can be summarized as "dis-

crimination under drive conditions." The preverbal organism comes to react to certain stimulus configurations that represent ingestibles or predators, and it is simply an extension of this process when an interpersonal vocabulary is introduced to refer to these configurations. The psychologist refers to both the preverbal and the verbal stages of this process as concept formation, and although many philosophers prefer to restrict the term to the verbal stage, there is no doubt that the verbal stage develops from the preverbal one, whether or not it becomes qualitatively different in the course of development. [See CONCEPT FORMATION.]

**Definition.** Given the pragmatic view of concept formation outlined above, it is very natural to abandon classical definition theory (in which a definition is a set of logically necessary and sufficient conditions for the use of a term) and to adopt instead an approach based on the notions of indicator and of indicator clusters. The rationale for such an approach may be stated as follows. Within a formal system, such as mathematics or an invented language, we can find examples of classical definitions, but seldom, if ever, can we find them within a natural language or the language of a developing science. The reason for this is simply that natural phenomena are extremely untidy, and a language for referring to them must be flexible enough to accommodate this untidiness if it is to be scientifically useful. For instance, the concepts "subject," "length," "response," "group," "aggression," and "temperature" cannot be given classical definitions (other than trivially circular ones). But we have no serious difficulty in learning or teaching these concepts, which we do by giving paradigm examples, contrasts, and explicitly approximate or conditional definitions. Any paradigm provides us, if it can be successfully analyzed, with a set of sufficient conditions for the application of the term, but it is not possible to insist that paradigms, let alone analysis of them, can never turn out to be erroneous. So the sense in which these conditions are sufficient is still not that of logical sufficiency. It is still more obvious that we often cannot specify any properties that are logically necessary for the application of the concept. We therefore adopt the notion of "weighted indicators," or criteria—that is, properties which are relevant to the application of the term, but in varying degrees. Some of the properties may, in some cases, be virtually necessary, some small groups of them may be sufficient; but these are special cases. The clustering of these properties in property space indicates an entity for which it is useful to have



a name, but there is typically considerable scatter and no sharp boundary to the cluster (the so-called open texture of concepts). The charm of a concise translation (that is, of a classical definition) is not often provided by the cluster approach, but it does preserve that most valuable feature of classical theory—a reliable verbal response that can be successfully taught. Contrasting and approximating, which are the other procedures involved in specification of meaning, depend on the psychological superiority of discrimination skills over absolute recognition skills and involve no new logical points.

At this point, it should be stressed that there are two elements in the theory of definition: the pragmatic and the literal. We have been talking about the limitations of the literal theory of definition as a kind of translation. But there is the independent question of how one can best—that is, most usefully—select the terms that one will mention in the definition, whether it is a translation or a cluster of indicators. A famous answer to this has often been taken as both literal and pragmatic, which makes it nonsensical instead of good pragmatic advice of a limited range of applicability. I refer to operationism (for which see especially Bridgman 1927). There are, of course, some areas where it is to be taken very seriously as advice, and psychology is one of them. (In theoretical physics, contrary to the common impression, it is extremely rare to seek or find operational definitions.) But if operationism is taken as requiring that definitions actually contain nothing but operations, then it is absurd, a category error—and, in addition, it is self-refuting, since the concept of an operation cannot be defined operationally. Definitions should therefore be of such a kind that we can apply some independently determinable criteria to decide when the defined term should be used (we can, if we wish, call the process of determining whether these criteria apply “operations”). But we cannot require that the only content of definitions should be operations, or we find ourselves caught in the dilemma of deciding when we have only one operation and, hence, only one concept . . . and so on.

**Classification.** The traditional requirements for classification also need to be relaxed. Instead of the Aristotelian requirement of exclusive and exhaustive subcategories, we must settle for minimized overlap and maximized coverage, balancing these desiderata against considerations of simplicity and applicability. The development of a scientific classification scheme is essentially a remapping of property space, using a new system of

coordinates or a new projection. Hence, assessment of the scheme's value very much depends, as in the map analogy, on the purposes for which it is intended (which may be as diverse as fast labeling, representation of a new theoretical insight, and easy recall) and very little on the idealized and context-free notion of “cutting Nature at the joints,” which was the goal of earlier taxonomists (and which still admirably expresses the feeling inspired by a successful classification scheme).

**Measurement.** The urge to measure is a kind of conditioned reflex among empirically trained scientists, and with some reason. Measurement requires—or, when developed, provides—an inter-subjectively reliable procedure for applying the potentially infinite descriptive vocabulary of number. For describing any continuous variable we need such a vocabulary, and for describing many discrete variables it is exceedingly useful because it creates the possibility of applying some mathematical apparatus. But the social scientist all too often assumes that scaling and measuring will make it likely that we can discover useful laws. In fact, many scales disguise the existence of regularities, and many others fail to reveal them. The descriptive utility of a scale must be considered independently of its theoretical fertility, for the two considerations may lead in different directions. The proliferation of scales, like the proliferation of statistical analyses, often merely clogs the channels of communication. There is no magic about numbers *per se*, and only hard thinking and good fortune, in combination, achieve a worthwhile new scale. The same is true of classification. The cluster concept is peculiarly suited to the needs of statistical scaling and to those of statistical inference in general. [See CLUSTERING; SCALING.]

**Evaluation.** The notion of measurement phases quite gradually into that of evaluation. “Pecking order,” “peer-group ranking,” and “intelligence quotient” can all, in certain contexts, appear as value-impregnated scales. This is often disputed by those who think “value” is a dirty word. Social scientists are peculiarly susceptible to the charms of the Weberian myth of a value-free social science, which rests on a one-sided interpretation of the distinction between facts and values. In its more defensible form this distinction amounts to the assertion that (a) evaluation depends on establishing certain facts; and (b) evaluation goes beyond the facts on which it depends. To this it can be objected that evaluation may well *result* in facts, albeit facts different from those with which it begins.

Thus *Consumer Reports* can often substantiate

evaluations such as "Brand X is the best dishwasher on the market today" on the basis of performance data plus data about the consumers' wants. And that evaluation, when suitably supported, simply states a fact about Brand X. It is one thing to be cautious about importing *dubious* value premises into an argument; it is another to suppose that all value premises or conclusions are dubious, "mere matters of opinion." It is not at all dubious that dishwashers should get the dishes clean, that clocks should keep good time, and so on. It may simply be a fact that Brand X of dishwasher excels at its task. Since the evaluation is nothing more than a combination of these facts, it is itself a fact, although it is also an evaluation.

Precisely this analysis applies to myriad issues in the social sciences—evaluation of the relative merits of protective tariff barriers and direct subsidies or of different forms of treatment for dysfunctional neurosis, to name only two examples. Of course, there are many other issues where only a relativized answer is possible, since legitimate differences of taste exist (that is, differences where no argument can establish the superiority of any particular taste). But this in no way distinguishes evaluation from theoretical interpretation. It will not do to say that "in principle" these differences can be resolved in the theoretical area when more facts are available, but not in the evaluation area. Not only are there very fundamental disagreements about the function of theory that make this improbable, but additional facts often reduce value disagreement. In any case, *relativized* answers, which are still value judgments, admit of no disagreement, and therefore do not involve subjectivity. It is simply naive to suppose that fields must, if they are to be regarded as objective, always contain answers to questions of the form "Which is the best X?" Neither mathematics nor engineering has this property, or any corresponding property with respect to "Which is the true solution to X?" where the conditions of X are incompletely specified.

As to the final defense of the distinction between fact and value—that a particular choice of ultimate values, or ends, cannot be proved true—three answers apply: (1) No such choice has to be proved true, only relevant. (2) No one has yet produced a demonstrably ultimate value, and it seems clear that the hierarchical model is as inappropriate here as it is in epistemology. (There are no ultimate, or basic, facts, only some that are more reliable than others; and all can—under pressure—be given support by appeal to others.) (3) In the case of moral values, where some need for a single

answer can be demonstrated, there is a perfectly good way of handling any moral disputes that are likely to arise in the social sciences. This is by appealing to the equality of *prima-facie* rights—the fundamental axiom of democracy. Justification of this is itself one of the most important tasks of the social sciences and requires contributions from game theory, political science, economics, sociology, and psychology, as well as from the obvious candidate, anthropology. It is the problem of deciding on the best system of sublegal mores—insofar, of course, as any answer is possible. But this *type* of problem is already familiar: it is an efficiency problem.

Indeed, the social sciences have for too long been reticent about pointing out the extent to which it is possible both to give answers to moral problems and to refute some of the current answers to them. For some reason, the sound point that there are many areas of morality where there cannot yet be any proof of the superiority of a particular answer is taken to imply that acting as if any answer were correct is defensible. But of course the only defensible position in such a situation is a compromise or, if possible, a suspension of action. In my opinion, the axiom of equal *prima-facie* rights can be shown to be superior to any alternative, and also to be capable of generating a complete moral system. It achieves both of these ends by being the only such axiom that can be satisfactorily supported, and of course it roughly coincides with the Golden Rule and other basic maxims (details of this line of argument will be found in Scriven 1966). At any rate, it is time that the possibility of this kind of "strategic ethics" was examined by all the strategists who have sprung up to look at other kinds of social interaction in terms of the language of games.

In summary, the scientific approach to evaluation necessarily involves both applied science (finding the best solution to a practical problem) and pure science (finding the best estimate, hypothesis, or experimental design). Moreover, the element of evaluation enters into much "straight" measurement—that of intelligence, creativity, and social acceptability, for instance. The reason for this is as simple as the reason for introducing new classifications: we have to order our knowledge if we are to use it efficiently. It follows that the most appropriate kind of order is determined by the use to which that knowledge is to be put. One might put the point paradoxically and say that all science is applied science. The application may be to efficient explanation, accurate prediction, or improvement of the world, but it is still—in one important



sense—an application. Nor can we assess (that is, evaluate) theories or instruments unless we have some analysis of their proposed use.

**Data analysis.** There are obviously limitations on the memory storage capacity of the human brain that would make it impotent in the face of even a modest amount of data about the path of a projectile or a planet, if these data were presented in discrete form. But if we can find some simple pattern that the data follow, especially if it is a pattern that we are familiar with from some other, preferably visual, experience, then we can handle it quite easily. (The preference for visual models is due to the vastly greater channel capacity of the visual modality; the notion of "handling" involves more than simple recovery, and even that involves more than storage capacity, since it includes look-up and read-out time.) This simple point can be taken as the start for a complete data-processing approach to the aspects of structural philosophy of science that remain to be discussed. It also pervades the parts already discussed, for the introduction of concepts is itself a device to cope with the "buzzing, blooming confusion" of sense experience.

**Generalizations and laws.** The simplest examples of the cognitive condensation described in the preceding section are laws and generalizations. They are so necessary that we are prepared to forgive them a very high degree of inaccuracy if only they will simplify the range of possibilities for us. Contrary to Karl Popper's view, which stresses the equivalence of a law to the denial of counterinstances, the present view suggests that we are rightly very unimpressed by counterinstances in this irregular world, and extremely thankful if we can find a regularity that provides a reasonable approximation over a good range of instances. This is as true in the physical sciences as in the social, for the laws of gases, the laws of motion, gravitation, radiation, and so on, are all merely approximations at best. There are occasional exceptions (for example, the third law of thermodynamics), but the important feature of a law is that it introduces a pattern where none was before. In this sense its truth is its utility, not its lack of error. A great deal has been made in the literature of the fact that a law will support counterfactual inferences while an accidental generalization will not (Nagel 1961, pp. 68–73; compare Strawson 1952, pp. 85, 197, 200). On the present view, all this means is that we think natural laws are true of continuing or pervasive phenomena, i.e., of the nature of things. All other attempts to give logical analyses of the concept of law—for instance,

the view that laws cannot contain reference to particular entities, or must support predictions—seem to me simply arbitrary.

**Explanation.** The task of explanation is the integration of "new" phenomena (whether subjectively or objectively new makes no difference) into the structure of knowledge. Typically, this consists in fitting these phenomena into a pattern with which we are already familiar. But explanation is not simply reduction to the familiar. Indeed, it may involve reducing the familiar to the unfamiliar, as Popper (1963, p. 63) has pointed out. But it must involve reducing the uncomprehended phenomenon to phenomena that are comprehended, and it is hardly surprising if this sometimes increases the sense of familiarity.

It has long been argued that the way in which explanation is performed in science always involves subsuming the phenomenon to be explained under a known law [see SCIENTIFIC EXPLANATION]. In view of the known deficiencies in our scientific laws, explanation, if this theory were acceptable, would in any case be a rather imprecise affair. But the restriction to laws is too confining, unless one treats the term "law" so loosely that it covers almost any statement except a pure particular. The real danger in this model of explanation—the "deductive" model, as it is usually called—is that it seems to support the view that explanations are closely related to predictions. In contrast, the "pattern" model of explanation, which I prefer, allows for retrospectively applicable patterns as well as for those which can be grasped in advance of their completion. It happens that this is very important in the behavioral sciences, since there is, and always will be, a real shortage of "two-way" laws (that is, laws that both predict and explain). This has typically been treated by social scientists as a sign of the immaturity of their subject; but in fact it is simply a sign of its nature and is very like the situation in the "messier" areas of the physical sciences, especially engineering. When a bridge fails, or a mob riots, or a patient commits suicide, there are often a number of possible explanations, of which only one "fits the facts." We thus have no difficulty in saying that this one is *the* explanation, although the facts available before the event were not enough to permit us to tell what would happen. Indeed, it may be that the chances in advance were very strongly against the actual outcome, and that we would therefore have been entitled to make a well-founded scientific prediction of the event's nonoccurrence. It is not helpful to suggest in such cases that we are explaining by deduction from a law. What sort of

law? That some factors *can* cause others? That *only* certain factors cause a certain effect? These are indeed general propositions, but they do not look much like laws of the usual kind, and they do not support any predictions at all. It is not in the least important whether or not we call them laws, but it is important to understand that we often support explanations by appealing to general claims that will not support predictions.

**Prediction.** On the other hand, it is often possible to predict on the basis of laws that do not explain. Any reliable correlation, however little we understand it, will give us a basis for predicting (that is, of course, provided there is temporal separation between the two correlated factors). To be able to predict, we need only to be able to infer that an event must occur at some conveniently remote time in the future. To explain, on the other hand, we have only to show *why* something occurred, not necessarily that it *had* to occur.

For a long time it was thought that the only way we could be sure to avoid *ad hoc* explanations was to require that (a) the grounds for the explanation be known to be true and (b) these same grounds imply only one possible outcome, the event or phenomenon to be explained. But there are many ways in which we can show that a particular explanation is correct in a particular case without having to show that the factors mentioned in the explanation would enable one to infer the explained event.

For example, we may demonstrate the presence and operation of a particular cause by exhibiting some clues that uniquely identify its *modus operandi*. Of course, we may still have reason to believe that there are certain other factors present which, in conjunction with the factors we have identified, would enable us to make a formal inference that the event we have explained was bound to occur. But in many cases we do not know what those other factors are, and we do not know the laws that would make the inference possible. In fact, our belief that we may still be able to make this inference merely indicates that we believe in determinism. It certainly does not prove that a good scientific explanation demonstrates the necessity of what it explains.

No doubt it is tempting to remark that a *complete* explanation would have this property, but even this approach is wrong. In the first place, there are many so-called complete scientific explanations that do not have it. Second, this approach recommends a use of the term "complete explanation" that has the unpromising property of identifying mere subsumption under a generaliza-

tion as complete explanation. But the genius of Freud, for instance, was not manifested by proving that the "explanation" of obsessive behavior in a particular neurotic was that this behavior was common in neurotics. Rather, he claimed to show a causal connection between certain other factors and obsessive behavior; and a causal connection is neither as strong nor as weak as a high correlation, it is simply quite different.

**Causes.** We are often able to identify causes in the social sciences without being able to give the laws and other factors in conjunction with which they operate to bring about their effects. The basic control group study is a case in point. It may give us excellent grounds for supposing, let us say, that a particular drug accelerates the subjects' rate of response, but it does not tell us what other factors are necessary conditions for this effect. Of course, we know that it is some set of the factors which are present, but we do not know which set. In more complex cases, even though a particular factor does not reliably lead to this effect, we may establish it as the cause simply by elimination of the alternative possible causes, either through their absence or through the presence of clues showing that, although present, they could not have been active this time. Of course, causal explanations are incomplete in the sense that there are always further interesting questions about the phenomenon that we would like answered. But this is also true even when we can subsume the explanation under a law: for instance, we can ask why the law is true. But causal explanations are often complete in the sense that they tell us exactly what we need to know, for our interest in a phenomenon is often very specific. Thus explanation is a context-dependent notion; explanations are devices for filling in our understanding, and the notion of *the* explanation makes sense only when there is a standard background of knowledge and understanding, as in the normal pedagogical development of an academic subject.

**Understanding.** The key notion behind that of explanation, and hence that of cause, is understanding. It must not be thought of purely as a subjective feeling; the feeling is only something associated with it. The condition of understanding itself is an objectively testable one; in fact, it regularly is tested by examinations, which are supposed to stress the detection of "real understanding," not just rote knowledge. Understanding is integrated, related knowledge; more generally — so that the definition will apply to the understanding possessed by machines, animals, and children — it is the capacity to produce the appropriate re-



sponse to novel stimuli within a certain range or field. It is a characteristic of tests for understanding that they present the student with new problems, and the possession of understanding is valuable just because it does embody this capacity to handle novelty.

Approaching it in this way, we see immediately that the question of how the brain can provide this capacity is of interest not only from a neurophysiological point of view but possibly also as a way of obtaining better insight into the nature of understanding itself. In fact, a remarkable overlap emerges between the requirements of efficient storage and understanding, as it is commonly conceived. Efficient storage, both to economize on storage capacity and to facilitate fast recovery and input, must use mapping, modeling, and simplifying procedures. But these are exactly what we use in the process of trying to understand phenomena, even in the more pragmatic process of trying to improve our classifications, descriptions, and predictions. Thus, to mention one example of how fruitful this conception is, the idea of simplicity, which has often been recognized as a criterion for satisfactory scientific theories, can be seen as equivalent to the familiar notion of economy, not as some mysterious aesthetic requirement. Nor is the whole aim of understanding to be seen as a concession to human weakness; for the most sophisticated and capacious computers we shall ever design will be hard pressed to store even part of the brain's data and will therefore have to employ procedures that economize on storage space. Putting the matter more strongly, a simple map stores an infinity of facts with very little complication in the read-out procedures, in comparison with list storage, and an enormous improvement in speed of readout (that is, of recall); it also handles a quantity of information that would be beyond the capacity of ordinary list storage. (The difference between the use of models and the use of mnemonics is that the model stores an indefinite number of facts, which the mnemonic does not, and the model stores them in a way that is more directly related to truth.) Models, from analogies to axiomatizations, are the key to understanding. [See VERSTEHEN.]

Experimentation. The use of applied science for the purpose of controlling reality has its counterpart within the methodology of science itself in the area of experimentation. Thus applied science is to pure science as experimentation is to pure observation. In each case the armory of relevant methods is importantly different: the "purer" subjects are more concerned with understanding and

description; the more applied ones, with causation and manipulation. Indeed, it does not appear possible to give an analysis of "cause" without some reference to manipulation, though a great deal can be said before coming to the irreducible element of intervention or action. [See CAUSATION.]

A number of topics have not been discussed here, although they are closely related. For example, the problem of the reduction of one science or theory to another is a combination of a problem about explanation with one about definition. I have also said very little, even implicitly, about the issue of free will versus determinism, and wish only to state that it now seems possible to reconcile the two positions, provided we first make a distinction between predictive and explanatory determinism. In conclusion, I should stress my conviction that there is very little in the social sciences that does not have a parallel in the physical sciences, but it has not been to these parallels that social scientists have turned for paradigms. They have turned instead to the absurdly oversimple paradigms of Newtonian mechanics and astronomy. Even there, the significant fact is that the problem of predicting the motion of pure point-masses, moving under the sole influence of the very simple force of gravity, passes from the realm where solutions and predictions are possible to the realm where solutions give way to (at best) explanations as soon as the number of bodies is increased from two to three. It is surely absurd to imagine that the forces acting on human beings are simpler than those involved in the "three-body problem." And if this is so, then we need a new "methodology of the complex domain," perhaps along the lines indicated above.

MICHAEL SCRIVEN

[Directly related are the entries CAUSATION; EXPERIMENTAL DESIGN; MODELS, MATHEMATICAL; MULTIVARIATE ANALYSIS; POSITIVISM; PREDICTION; PROBABILITY; SCIENCE, article on THE HISTORY OF SCIENCE; SCIENTIFIC EXPLANATION; STATISTICS, article on THE FIELD; SYSTEMS ANALYSIS, article on GENERAL SYSTEMS THEORY; VALUES, VERSTEHEN; and in the biographies of COHEN; GALTON; HUME; KOYRÉ; PEARSON; PEIRCE; SCHLICK; WEBER, MAX; YULE.]

#### BIBLIOGRAPHY

*Contemporary philosophers of science can be roughly classified as either post-positivists (who form the majority) or linguistic analysts. For examples of work by leading post-positivists, see Popper 1934, 1957; Carnap 1950, Reichenbach 1951; Hempel 1952; 1965; Braithwaite 1953; Nagel 1961; Feyerabend 1962; Kemeny & Snell 1962; Suppes & Zinnes 1963. There are many important topics on which*

these authors disagree, but it seems fair to say that among the predecessors they would be prepared to acknowledge are such founders of modern positivism as Karl Pearson, Moritz Schlick, and Ernst Mach. Other outstanding earlier writers who, although somewhat untypical of the field, have contributed to the philosophy of science in general, not just the post-positivist branch of it, include C. S. Peirce, Max Weber, N. R. Campbell, P. W. Bridgman, and Harold Jeffreys. Works by philosophers of science who use at least some of the methods of the linguistic analysts, even if they would not all classify themselves as such, include, in the study of the physical sciences, Toulmin 1953; Polanyi 1958; in the study of the social sciences, Dray 1957; Winch 1958; Kaplan 1964; Scriven 1966; and, in the field of measurement and probability, Ellis 1966; Hacking 1965. Certain trends in the history of science, particularly in the emphasis, developed in Butterfield 1950 and Kuhn 1962, and by Alexandre Koyré, on autonomous conceptual schemes as units of progress, have influenced the philosophy of science. These trends have in turn been influenced by philosophical works such as Feyerabend 1962; Hesse 1963. Among the extremely valuable contributions to the philosophy of science made by practicing social scientists, special note should be taken of Simon 1947-1956; Skinner 1953; Meehl 1954; Stevens 1958; Chomsky 1966. Of the general treatises available in 1966, only McEwen 1963 and Kaplan 1964 are concerned principally with the social sciences, although Nagel 1961 and Hempel 1965 deal with them to some extent. Lazarsfeld 1954; Natanson 1963; Braybrooke 1966 are interesting collections that focus exclusively on the social sciences. As an introduction to the philosophical problems of the social sciences, Kaplan 1964 is to be recommended to layman and expert alike, although the work takes an approach somewhat different from that presented in this article.

- BRAITHWAITE, R. B. 1953 *Scientific Explanation: A Study of the Function of Theory, Probability and Law in Science*. Cambridge Univ. Press. → A paperback edition was published in 1960 by Harper.
- BRAYBROOKE, DAVID (editor) 1966 *Philosophical Problems of the Social Sciences*. New York: Macmillan.
- BRIDGMAN, P. W. (1927) 1946 *The Logic of Modern Physics*. New York: Macmillan.
- BUTTERFIELD, HERBERT (1950) 1957 *The Origins of Modern Science, 1300-1800*. 2d ed., rev. New York: Macmillan. → A paperback edition was published in 1962 by Collier.
- CAMPBELL, NORMAN R. (1920) 1957 *Foundations of Science: The Philosophy of Theory and Experiment*. New York: Dover. → First published as *Physics: The Elements*.
- CARNAP, RUDOLF (1950) 1962 *The Logical Foundations of Probability*. 2d ed. Univ. of Chicago Press.
- CHOMSKY, NOAM 1966 *Cartesian Linguistics: A Chapter in the History of Rationalist Thought*. New York: Harper.
- DRAY, WILLIAM (1957) 1964 *Laws and Explanation in History*. Oxford Univ. Press.
- ELLIS, BRIAN 1966 *Basic Concepts of Measurement*. Cambridge Univ. Press.
- FEIGL, HERBERT; and SCRIVEN, MICHAEL (editors) 1956 *The Foundations of Science and the Concepts of Psychology and Psychoanalysis*. Minnesota Studies in the Philosophy of Science, Vol. 1. Minneapolis: Univ. of Minnesota Press.
- FEYERABEND, P. K. 1962 *Explanation, Reduction and Empiricism*. Pages 28-97 in Herbert Feigl and Grover Maxwell (editors), *Scientific Explanation, Space, and Time*. Minnesota Studies in the Philosophy of Science, Vol. 3. Minneapolis: Univ. of Minnesota Press.
- HACKING, IAN 1965 *Logic of Statistical Inference*. Cambridge Univ. Press.
- HEMPEL, CARL G. 1952 *Fundamentals of Concept Formation in Empirical Science*. Volume 2, number 7, in *International Encyclopedia of Unified Science*. Univ. of Chicago Press.
- HEMPEL, CARL G. 1965 *Aspects of Scientific Explanation, and Other Essays in the Philosophy of Science*. New York: Free Press.
- HESSE, MARY B. 1963 *Models and Analogies in Science*. London: Sheed & Ward.
- JEFFREYS, HAROLD (1931) 1957 *Scientific Inference*. 2d ed. Cambridge Univ. Press.
- KAPLAN, ABRAHAM 1964 *The Conduct of Inquiry: Methodology for Behavioral Science*. San Francisco: Chandler.
- KEMENY, JOHN G.; and SNELL, J. LAURIE 1962 *Mathematical Models in the Social Sciences*. Boston: Ginn.
- KUHN, THOMAS S. 1962 *The Structure of Scientific Revolutions*. Univ. of Chicago Press. → A paperback edition was published in 1964.
- LAZARSFELD, PAUL F. (editor) (1954) 1955 *Mathematical Thinking in the Social Sciences*. 2d ed., rev. Glencoe, Ill.: Free Press.
- MC EWEN, WILLIAM P. 1963 *The Problem of Social-scientific Knowledge*. Totowa, N.J.: Bedminster.
- MEEHL, PAUL E. (1954) 1956 *Clinical Versus Statistical Prediction: A Theoretical Analysis and a Review of the Evidence*. Minneapolis: Univ. of Minnesota Press.
- NAGEL, ERNEST 1961 *The Structure of Science: Problems in the Logic of Scientific Explanation*. New York: Harcourt.
- NATANSON, MAURICE (editor) 1963 *Philosophy of the Social Sciences: A Reader*. New York: Random House.
- POLANYI, MICHAEL 1958 *Personal Knowledge: Towards a Post-critical Philosophy*. Univ. of Chicago Press.
- POPPER, KARL R. (1934) 1959 *The Logic of Scientific Discovery*. New York: Basic Books. → First published as *Logik der Forschung*.
- POPPER, KARL R. 1957 *The Poverty of Historicism*. Boston: Beacon.
- POPPER, KARL R. 1963 *Conjectures and Refutations: The Growth of Scientific Knowledge*. New York: Basic Books; London: Routledge.
- REICHENBACH, HANS (1951) 1959 *Probability Methods in Social Science*. Pages 121-128 in Daniel Lerner and Harold D. Lasswell (editors), *The Policy Sciences: Recent Developments in Scope and Method*. Stanford Univ. Press.
- SCRIVEN, MICHAEL 1956 *A Possible Distinction Between Traditional Scientific Disciplines and the Study of Human Behavior*. Pages 88-130 in Herbert Feigl and Michael Scriven (editors), *The Foundations of Science and the Concepts of Psychology and Psychoanalysis*. Minnesota Studies in the Philosophy of Science, Vol. 1. Minneapolis: Univ. of Minnesota Press.
- SCRIVEN, MICHAEL 1959 *Truisms as the Grounds for Historical Explanations*. Pages 443-475 in Patrick L. Gardiner (editor), *Theories of History*. Glencoe, Ill.: Free Press.
- SCRIVEN, MICHAEL 1962 *Explanations, Predictions and Laws*. Pages 170-230 in Herbert Feigl and Grover Maxwell (editors), *Scientific Explanation, Space, and Time*. Minnesota Studies in the Philosophy of Science, Vol. 3. Minneapolis: Univ. of Minnesota Press.
- SCRIVEN, MICHAEL 1965 *An Essential Unpredictability in Human Behavior*. Pages 411-425 in Benjamin B.



- Wolman and Ernest Nagel (editors), *Scientific Psychology: Principles and Approaches*. New York: Basic Books.
- SCRIVEN, MICHAEL 1966 *Primary Philosophy*. New York: McGraw-Hill.
- SIMON, HERBERT A. (1947-1956) 1957 *Models of Man; Social and Rational: Mathematical Essays on Rational Human Behavior in a Social Setting*. New York: Wiley.
- SKINNER, B. F. (1953) 1964 *Science and Human Behavior*. New York: Macmillan.
- STEVENS, S. S. 1958 Problems and Methods of Psychophysics. *Psychological Bulletin* 55:177-196.
- STRAWSON, P. F. 1952 *Introduction to Logical Theory*. London: Methuen; New York: Wiley.
- SUPPES, PATRICK; and ZINNES, JOSEPH L. 1963 Basic Measurement Theory. Volume 1, pages 1-76 in R. Duncan Luce, Robert R. Bush, and Eugene Galanter (editors), *Handbook of Mathematical Psychology*. New York: Wiley.
- TOULMIN, STEPHEN 1953 *The Philosophy of Science: An Introduction*. London: Hutchinson's University Library. → A paperback edition was published in 1960 by Harper.
- WINCH, PETER (1958) 1963 *The Idea of a Social Science and Its Relation to Philosophy*. London: Routledge; New York: Humanities.

## III

## THE SOCIOLOGY OF SCIENCE

If the sociology of knowledge is defined as the part of sociology that studies the nature of and relations between different types of idea systems, on the one side, and the relations between these idea systems and a variety of institutional (or social-structural) and personality factors, on the other, then the sociology of science is one part of the sociology of knowledge. It is the part that specializes in defining the nature of scientific ideas and in describing their relations both to other kinds of ideas (e.g., ideological, philosophical, aesthetic, religious) and to various institutional and personality factors. Parsons (1951, especially chapter 8) has given us what is still the best analytic definition of idea systems in general and their several specialized subtypes, although further analysis and operational specification of his classification are necessary.

The sociology of science, as is the case with all sociology, general or special, is primarily interested in the construction of a set of highly generalized, systematic, and relatively exhaustive concepts and propositions of relationship. In this enterprise it uses data from all historical periods and all cultures, since its main concern is not with history as such, but with establishing sociological concepts and propositions. The history of science, although it always should use such explicit sociological concepts and propositions, often does not, preferring

its traditional, less examined, and frequently common-sense ways of treating its materials. In either case, the history of science, both advertently and inadvertently, may be productive of materials or even of new concepts and propositions for the sociology of science, but this is not its essential task. Thus, the sociology of science and the history of science overlap but do not coincide.

The sociology of science, finally, is interested both in fundamental scientific ideas themselves and in the application of these fundamental ideas, or of more empirical ideas, to technology. In its study of technology, again, the sociology of science uses both historical and contemporary data, drawn from a variety of cultures, regardless of the original purpose for which these data were collected, so long as they serve its primary goal of constructing a system of analytic concepts and propositions.

The social nature of science. One of the basic working assumptions of structural-functional theory in sociology is that man's behavior in society is a response to certain functional problems that he confronts in his nonsocial and social environments. On this assumption, sociology sees man's scientific behavior as his response to the functional problem created by his need to have adequate adjustive knowledge of the physical, biological, and social aspects of the empirical world. Such knowledge is indispensable for some form of adjustment to that world in its three essential and different aspects. Everywhere, as archeology, ethnology, history, and sociology have severally demonstrated, man in society has a certain amount of this scientific knowledge. The amount, of course, varies a great deal among societies. In other words, science, as Malinowski recognized (1916-1941), is a matter of degree, with some societies having only a relatively small amount and others having a great deal more. This, indeed, sets some of the essential problems for the sociology of science. For it is not only the universal occurrence of at least a small amount of scientific knowledge, but also the patterns of variation in the degree of development of such knowledge in different societies, and variations at different stages in the evolution of a single society, that the sociology of science has undertaken to explain.

In order to explain the degree of development of science considered as a complex whole, it is necessary to distinguish its component cultural, social, and psychological parts, for each of these can develop at somewhat different rates in different social situations. It is useful to define at least four general components of science as a whole,

each of which has a measure of autonomy: substantive scientific ideas; scientific methodology, including both ideas and instruments; scientific roles; and motivational and reward systems for scientific roles. Let us consider, again because this sets important problems that the sociology of science has been trying to explain, how each of the components varies along certain dimensions.

Substantive scientific ideas vary along three dimensions: of generality, or abstractness, as it is also often called; of systematization; and of exhaustiveness for the relevant aspects of the phenomena. We may say that the more abstract, the more systematic, and the more nearly exhaustive a set of substantive scientific ideas is, the greater the degree of its scientific development. Among the interesting problems that the sociology of science tries to explain is the question of why the physical sciences on the whole have developed more rapidly in all three respects than have the biological sciences and why these, in turn, have developed more rapidly than the social sciences. Within each of these broad sectors of science, moreover, further variation in the degree of development of specialized subspecies poses similar problems for explanation. For example, the more rapid recent development of economics than of political science or sociology can be explained, in part, by the greater accessibility of data on price phenomena than of data on people's political and social norms and actual behavior. This greater accessibility is due, in turn, and again in part, to the more pressing felt needs that governments and business firms have for price data.

Knowledge about the ideas and instruments (or technology) that make up scientific methodology also varies in its degree of development. It is clear that modern science has made great advances over earlier science in the sophistication of its knowledge about such essential methodological matters as the character of concepts and classifications, the logic of comparison and inference, and the functions of contrived and natural experimentation. Modern science is also at a great advantage over earlier science in the degree of its access to ordering and measuring facilities. Mathematics is, of course, the most important of such facilities, and its development has been of very great moment in the closely related development of all the sciences, although, again, to varying degrees among the several sciences, for interesting sociological reasons. In addition to mathematics, such instruments as thermometers, high-speed electronic computers, and a thousand others have contributed significantly to the advances of the sciences. Both

the development of these measuring facilities and their adoption by the several sciences are important focuses for research in the sociology of science. Since they may develop sometimes as a direct result of changes in scientific ideas, but also sometimes independently, methodological ideas and instruments have various types of relationships with substantive scientific ideas. When they develop to some extent independently, as in the cases of the telescope or the computer, for example, methodological ideas and instruments may at certain times play the leading part in the development of a whole sector or subsector of science. Or again, it is an important sociological fact about the prospects of the social sciences that their ability to borrow a great deal of the methodological sophistication developed in the other sectors of science is a key asset for their own development.

There is also considerable variation among societies in the degree of differentiation and specialization of social roles for engaging in the many types of scientific behavior, in the proportion of members of the society engaged in these roles, in the kinds and amounts of support given to these roles, and in the degree of institutionalization of scientific roles. The sociology of science has been much concerned with these essential problems of the social organization or role component of science, so much so that we shall take them up below, in a separate and detailed discussion.

Finally, motivational and reward systems for scientific roles also vary among and within societies. They differ in the type of rewards, in the strength of motivation, and in the degree of diffusion of adequate incentives throughout the population as a whole. Again, we shall consider this matter more fully below.

### The development of science

Viewed in the long perspective, science shows a history of continuity and cumulation from the earliest prehistory of man to the present. However, this history has been marked by quite different rates of development in different times and places. It is this unevenness in its progress, viewed in terms of shorter perspectives, that calls for explanation by the sociology of science. As the sociology of science has itself made progress, a few basic principles of explanation have come to appear sound. One is the principle that no single cultural, social, or psychological factor, such as religion or economic forces, can account for the growth of science as a whole or any of its several components or subspecies. Another such principle



is that not even any single combination of factors, such as the religious and the economic, or the political and the educational, is sufficient for the tasks of the sociology of science. This is not to say that there are not some combinations of factors more favorable than others for certain kinds of specific development in science, as our discussion of the Protestant ethic will illustrate. There definitely are such favorable combinations, and the sociology of science has improved its ability to discover them by accepting the principle that certain types and rates of development may come from one set of factors (or one set of values of these factors taken as variables) and that other types and rates may come from a different set of factors (or a different set of variables-values).

The following discussion will consider some of the major factors that affect the development of science. They exert their influence on science *always* in combination, and the relative weight of the influence of each may vary in different specific instances.

**Structural and cultural differentiation.** The greater the degree of structural differentiation in a society, that is, of the specialization of roles between the major institutional sectors and within them as well, the more favorable the situation for the development of science. Greater structural differentiation provides the necessary situation not only for a variety of highly specialized scientific roles but also for their highly specialized ancillary and supporting roles in other institutional sectors.

Besides role differentiation, differentiation within culture, that is, in the various systems of ideas in society, is also favorable to the development of science. For example, as Charles Gillispie (1960) has shown, the development of what he calls "the edge of objectivity" of scientific ideas has increased as these ideas have progressively been differentiated from moral ideas about man's place in the universe. The more that values, ideologies, aesthetic ideas, and philosophical ideas are distinguished both from one another and from substantive scientific ideas, the easier it is to see the special problems of each and to develop each and all of them.

**Value systems.** Certain values which the modern world tends to take for granted, but which have not existed in many past societies, certainly in nothing like the same degree, are much more favorable to the development of science than are their opposites. The high value the modern world puts on rationality as against traditionalism, on this-worldly activities as against other-worldly activities, on libertarianism as against authoritarianism, on active striving as against passive adapta-

tion to the world, and on equality as against inequality—all these values support the development of the several components of science. Sometimes the support is direct, as in the case of the values of rationality and this-worldly interest, values which are especially powerful in combination, as they are in the modern world. Libertarianism is essential for academic freedom, which is one kind of important foundation for scientific progress. Sometimes the support from values is indirect, as when the value of equality increases the amount of social mobility and thus helps to select better talent for scientific roles.

**Instrumental needs.** Scientific knowledge is power, that is, power to adjust more or less satisfactorily to the nonsocial environment and to the internal and external social environment. Some scientific discovery and technological innovation, then, is a response to immediate instrumental needs for adjustment to what are defined as dangerous and harsh environments. Certainly this must have been an especially large influence on the development of science in the early stages of human society, when every bit of scientific knowledge was of great instrumental importance in a threatening and severe physical and social environment. But even powerful modern industrial societies respond strongly to their instrumental needs for science. Whatever their values in regard to science, they feel an urgent need to use it to strengthen their national defense, promote industrial and agricultural growth, and improve the health of their populations. Some scientists deplore this "exploitation" of science, that is, this encouragement of science "not for science's own sake." But the instrumental needs of even already powerful societies are defined by their citizens as more important under some circumstances than the value considerations that are preferred by scientists and by those who share the values that support science directly. Finally, the nonindustrial or underdeveloped societies of the present also push the acquisition of science for urgent instrumental needs, to cope with "the revolution of rising expectations" in their populations.

**Economic factors.** A variety of economic structures, needs, and resources, often combined with one or more other social factors, have had very large direct and indirect influences on the development of science. There is no blinking this fact, no matter how much some scientists may think it stains the "purity" of science and regardless of whether one thinks some Marxist sociologists of science have exaggerated it. Only a few of the myriad effects of economic activities on science

can be given in illustration here. As Merton has shown (1936; 1939), in the case of seventeenth-century England, for example, the growing maritime interests, both commercial and naval shipping, were in need of more reliable techniques of navigation. The existing techniques included neither good chronometers nor any sure way of determining longitude at sea. These economic and military-political needs were a direct stimulus, through prizes offered for discoveries and inventions, to basic discoveries in astronomy and on the nature of the spring. Or again, earlier, in sixteenth-century Europe, economic needs for more powerful methods of pumping water out of mines than were provided by the existing Archimedean screw pumps led to work by Evangelista Torricelli and Vincenzo Viviani on the relations between atmospheric pressure and water level. On the basis of this work, successful suction pumps for industrial use were constructed. But this scientific work also led to the invention of the barometer by Torricelli, to the creation of a vacuum by Otto von Guericke, and to a whole host of scientific experiments and results with the so-called "air pump" by Robert Boyle and his colleagues in the newly founded Royal Society. This example shows how economic needs, science, and technology affect each other in complex and often mutually beneficial ways.

In the modern world, both industrial organizations and governments in all societies have provided many kinds of support for science in order to pursue economic interests and needs. Direct support exists in the form of governmental and industrial research laboratories; indirect support is given through tax provisions and other subsidies to both the universities and industry. One of the largest areas of governmental response to economic needs has been agricultural research, which is largely biological research but also includes the physical and social sciences. It should be noted that both "capitalist" and "socialist" societies and both "planning" and "nonplanning" societies have supported science through their responsiveness to the economic interests of special groups in the society and of society as a whole. The desirability of using science to meet economic needs and increase economic resources is no longer anywhere in question among the societies of the modern world.

**Political structures and needs.** Political and economic factors have often tended to be closely combined in their effects on science. Of course, the political factor has also had a great influence on science separately, for example, because of the military needs that governments have had. After the recent history of the part that governments have

played in the development of atomic science for politico-military needs, this point hardly needs further illustration. But this connection between politico-military needs and science is far from new. Governments have always sought to increase military firepower through the development of science. Research on the nature of metals, on the causes of chemical reactions and explosions, and on the mathematics of the curves of ballistic missiles has been spurred by military needs.

**Religion.** Values and religion are closely connected in every society. A religion that supports modern values, which in turn favor science, is, of course, a strong support for the development of science. This specific connection is what Merton has sought to demonstrate in his analysis of the relations between the complex of values and religious beliefs that constitute the Protestant ethic, on the one hand, and the great development of seventeenth-century English science (1936; 1939). The values and beliefs of the Protestant ethic have now been secularized and diffused to many social milieus other than the ones in which they originally arose. In their secular form they are even more powerful supports for the development of science. The relations between a religion and science are always, it should be carefully noted, complex and at partial cross purposes. Thus, although the general beliefs and values of the liberal forms of Protestant Christianity have been favorable to the development of science, even these liberal elements have sometimes protested against specific scientific discoveries that have apparently contradicted basic doctrines of their own. Thus, as Gillispie (1951) has shown, some liberal Protestants who still held to the Biblical conceptions of miraculous floods and other catastrophes were opposed to the notion of geological uniformitarianism when this new scientific idea was first developed in the early nineteenth century. And, a little later, as Dupree (1959) and Lurie (1960) have suggested, there was the same kind of resistance for a while to Darwin's discovery of uniformities in the development of man and the other animal species. In analyzing the relations between religion and science, the sociology of science treats them both as complex phenomena and looks for specific connections between their specific components.

**The educational system.** The maintenance and development of science is facilitated by an educational system that is sufficiently specialized for, and sympathetic to, the growing science of its time. For example, as we may learn from Ben-David (1960), new educational institutions developed for that purpose played an essential role in



the development of French science during the early nineteenth century and in Germany a little later. All the modernizing countries of recent times, from Russia to the small and weak African countries, have greatly strengthened their educational systems at all levels because of their desire to improve their science. Before modern times, the universities were not the home of scientific development, which they have since become. As late as "the Great Instauration" of science in seventeenth-century England, the universities were still indifferent or hostile to science, but a variety of other extremely favorable social conditions helped science to flourish. This hostility continued for some time. Now, of course, a university that is hostile or indifferent to science can exist only in a society, or in one of its parts, that is content to remain a scientific backwater.

**Social stratification system.** A system of social stratification that emphasizes the value of equality and that, in fact, realizes a high degree of social mobility for the members of a society seems to be more favorable for the development of science than is the opposite, or closed, type of system. By providing a greater degree of equality for talent, wherever in the society it originates, the open type of system can provide greater resources of talent for manning scientific posts. Of course, if scientific roles are not highly valued, rising talent will go elsewhere, but the two factors in combination—a high value on science and the openness of the channels of social mobility—are very favorable to the development of science. In the modern world, both of these conditions have been realized in ever greater measure in the older industrial societies. And the newly modernizing societies, as Dedijer (1961) has described them, are making great efforts to bring them into being almost overnight.

### Differentiation of the scientific role

In relatively undifferentiated societies, past or contemporary, there are few roles for full-time "intellectuals," as we may roughly characterize all those whose special function is to deal in some kind of idea system. However, as societies become less simple and more structurally differentiated, the role of intellectual enlarges and is occupied by at least a few individuals on a full-time basis and with support from a variety of sources: a religious organization, the political magnates, or inherited wealth. Certainly in Greece by the fifth century B.C., the role of intellectual had emerged and was occupied with great distinction by men who have influenced all subsequent thought—men like Plato and Aristotle. It should be noted, however, that the

ideas with which such intellectuals dealt were relatively much less differentiated than ideas are in the modern world. On the whole, the intellectual still deals alike with religious, scientific, ethical, political, and philosophical ideas; indeed, the term "philosophy" covers all knowledge, all types of ideas. But it is not until the beginnings of the modern world that this further process of differentiation occurs, that is, differentiation between types of ideas, and makes way for the specialized roles dealing wholly or primarily with scientific ideas. This differentiation of the scientific role from other intellectual roles was slow in occurring, even in the modern world. In seventeenth-century England, the new science was still called "the new philosophy." And the oldest scientific society in the United States, founded in 1743 by Benjamin Franklin, among others, was called the American Philosophical Society. So little differentiated was the specialized role of scientist and so few were those who occupied this role, by whatever name and with whatever means of support, that it was not until the nineteenth century that the very term "scientist" was coined (by the Rev. William Whewell, in England in the 1840s).

The process of differentiation of the role of scientist went slowly until the middle of the nineteenth century because of the difficulty of finding sources of support for it in the securely established social organizations and arrangements. Beginning in the sixteenth century, newly founded scientific societies provided various necessary facilities and supports for scientists, but only a handful of full-time jobs. Governments did something more, but still there were very few full-time positions in government museums or research organizations. Some distinguished scientists, men like Franklin and Boyle, supported themselves on inherited or earned wealth. Others worked at nonscientific jobs as well, or alternated between scientific and other activities; such a scientist was Antoine Lavoisier. Very little was done by the universities and colleges of the time, not until the nineteenth century, and in most places only in the latter half of that century, did they open their doors wide to scientists.

Full institutionalization was achieved, then, when universities, various governmental organizations, and many industrial firms recognized their great need for science, and established regular and permanent roles and careers for scientists. In addition, a variety of special research institutes, endowed by foundations, trade associations, and other special interest groups, or sometimes operated by private individuals for profit, provided further and ever more specialized jobs for scientists.

The necessity and the legitimacy of the scientific role were increasingly acknowledged.

One rough measure of the establishment of the scientific role in the modern world is the quantitative increase in the numbers of those who occupy the role. As Derek Price (1963) has shown, during the last 300 to 400 years there has been an exponential growth rate in the number of scientists in the modern world. Starting from very small numbers, to be sure, the rate of growth is such that there has been a doubling of the number of scientists in something like every 10 to 15 years. Because of the nature of exponential growth rates, the recent increase in the number of scientists has been especially large. Price estimates that about 90 per cent of all the scientists who have ever lived are still alive today. Given this growth rate, as well as other reasons, it is small wonder that the place of science in the modern world is not yet as settled as many would like.

There is, unfortunately, no room here to treat some other aspects of the social organization of science. Such matters as patterns of authority, of collaboration, and of careers in scientific work have recently been well studied by Glaser (1964), Hagstrom (1965), and Zuckerman (1965), among others.

### Motivations and rewards for scientists

The mere existence of social roles is not enough for their full institutionalization. Adequate and legitimate motivations in those who occupy the roles, and adequate and legitimate rewards from those who support the roles, must also exist. On the whole, in the modern world, and especially in the societies that modernized early, these motivations and rewards exist in sufficient measure.

What little evidence we have on the motives of working scientists indicates that, as in all other social roles, a variety of motives in different combinations is at work. For example, because the role of scientist now provides considerable stability, security, and prestige, many scientists are motivated in some measure to achieve these goals. Also, because specialization in basic scientific research provides considerable autonomy to the working scientist, many are motivated to achieve the independence and self-control that the role makes possible. In future social-psychological research on motivation for the scientific role, it will be desirable to keep in mind both the multiplicity of motives that can engage the competent scientist and the way in which the nature of his role structures certain typical combinations of motives. It would be interesting to establish the precise role

of much-vaunted "curiosity" in the motivation of scientists. And it would also be desirable to use a typology of different scientific roles, along the lines suggested by Znaniecki (1940), such as theorist, experimenter, synthesizer, and similar functional types, to see if various typical combinations of motives exist in these different scientific specialties.

As for rewards, the scientific role is now given satisfactory measures of such rewards as security, prestige, and money income to attract its fair share of the talented members of modern societies. As compared with those in business occupations, scientists tend to be rewarded with relatively more prestige for professional standing and relatively less money income. Among scientists themselves, and among the lay public to a lesser extent, there has developed a very elaborate set of symbolic rewards for differential prestige bestowed for differential achievement in science. The awarding of titles, prizes, medals, offices, and eponymous distinctions symbolizes the existing hierarchy of differential prestige. A Nobel Prize in any of the four scientific fields in which it is granted is only the best-known of the symbols of scientific achievement.

The reward of prestige in science is given primarily for originality in scientific discovery, but, since it is sometimes difficult to appraise the degree of originality, sheer productivity may carry off the honors. Like other men in other roles, scientists are deeply concerned about the just distribution of rewards for their activities. Despite the norm of humility that prevails among them and that does influence their behavior, they are motivated to achieve credit for being original, just as businessmen are motivated to receive credit for making a profit. As a result of their concern, as Merton (1957) has so well shown, scientists take pride in priority of discovery and often engage in bitter quarrels over claims to priority. In sum, the role of the scientist is subject, as are all social roles, to a structured set of motivations and rewards, some of which are similar to those in other roles and some of which are different. The scientist is in no sense a "selfless" creature above and beyond the influences of his social role.

During the last decade there has occurred the beginning of systematic research on what is called "the image" of the scientist that the public at large and various segments of the whole society, such as youth or the scientists themselves, actually hold. A favorable image of a social role is, of course, an important reward for those who fill it. The evidence accumulated by this research shows, on the whole, that the image of the scientist is a favorable one.



But as Mead and Métraux (1957–1958) and Beardslee and O'Dowd (1961) have shown in their research, there are some negative tones in the picture—a fear and dislike for some of the characteristics and consequences of the scientific role. In short, there is some ambivalence toward scientists. In certain quarters, this admixture of negative feelings has been much deplored, despite the fact that such feelings probably exist for every social role. Because all social roles probably work at least some negative consequences for some people in some situations, it is utopian to expect a wholly positive image for scientists or for those who occupy any other social-role category.

### Communication among scientists

Since one essential ingredient in the development of science is the combination of already existing ideas, effective communication among scientists is an indispensable part of scientific activities. One of the important social inventions in the early modern period in science was the creation of local, national, and international societies and journals as means for the speedier and more general communication of scientific work within the community of scientists. In order to “keep up” in any branch of science, that is, to learn about the new ideas he can use to discover still further new ideas, the working scientist now has to spend a valuable part of his time with the professional journals, carefully studying a few and scanning many others. In addition to reading the journals, scientists communicate at the meetings of professional associations, both formally and informally. Here, too, as science has developed, there has been a proliferation of specialized groups to match the specialized journals and the specialized activities they report. And, finally, through informal visits, letters, telephone calls, and preprint mimeographed materials, scientists strive to maintain the effective communication without which their activities would be slowed down and even stopped.

In his description of the exponential growth rates for different aspects of science, Price (1963) has also shown that there has been a doubling in the number of scientific journals every 10 to 15 years over the last three or four centuries. Because of this large increase, much of it necessary because of the growth of new scientific specialties, there has been some concern among scientists about the possibility of an excess of information; for some scientists, just “keeping up” has become ever more difficult. In response to this felt difficulty, abstracting services have been created and have multiplied, but still the problem of excessive and inefficient

communication is felt to persist. At the present, therefore, various groups of scientists are trying to codify and computerize the processes of what is called “information retrieval” in science. So far they have not had very great success, partly for reasons suggested below. [See INFORMATION STORAGE AND RETRIEVAL.]

The structure and functions of communications processes among scientists present an obvious set of problems for the sociology of science—one on which much remains to be done. One interesting suggestion from general sociological ideas and from the research that has been done by Herbert Menzel (1959) is that scientists themselves sometimes think in terms of too rational a conception of the communication process; that is, they may be expecting too much from the journals and the formal meetings. In addition to these formal, manifest, and planned communication processes in science, there are the informal, latent, and unplanned ones. Scientists cannot always know precisely what they want and merely push a computer button to get it. Often, through “milling around” at meetings, through chance visits, through indirect channels, they get essential information which they can recognize as essential only when they get it. As a journey into emergent novelty, science must use both planned and unplanned patterns of communication. One of the newer and more interesting focuses of research in the sociology of science is the question of the function of each pattern of communication for different scientific needs and the distribution of these patterns among the different social situations in which scientists find themselves. [See DIFFUSION, *article on* INTERPERSONAL INFLUENCE.]

### The processes of scientific discovery

In contrast with an individualistic, “heroic” conception of the processes of scientific discovery that has prevailed in some quarters, though less now than formerly, the sociology of science has sought to redress the balance of our understanding of these processes by demonstrating that they have essential social components. Starting with the fact of the cumulative nature of science, sociologists have pointed out that the established body of scientific ideas and methodologies at any given time itself has the most important influence on setting problems for scientists to solve and on providing leads for their solution. This determining effect of the established ideas and methods is the source of the innumerable examples of the pattern of independent multiple discovery in science. Given the prerequisites of a discovery in the established body

of science, it is almost inevitable, as Ogburn (1922) and Merton (1961) have argued and demonstrated from the history of science, that independent multiples will occur. Some discoveries, of course, break somewhat more sharply than others with the fundamental notions of the established science; they have more emergent novelty. These are what Kuhn (1962) has called "scientific revolutions." However, even these are far from *ex nihilo*. Although cumulative in important measure, they also set new directions for the relevant fields and specialties in science.

Recent theory and research in the sociology of science have also qualified the older picture of the processes of scientific discovery as based entirely on foresight, planning, rationality, and ready acceptance. These characteristics certainly can be seen in large measure in most discoveries. But, in addition, in many discoveries there is an admixture of the unplanned, the nonrational or irrational, and the obstructive, contributed by the discoverer himself or by others. These characteristics manifest themselves in the pattern of serendipity in discovery and in the pattern of resistance by scientists themselves to certain scientific discoveries. The serendipity pattern occurs, and in actual research it occurs very often, when the researcher comes by happy chance on something he was not looking for, that is, some anomaly that presents him with the unexpected opportunity to change his preconceptions about his research and make a new discovery (Barber 1952, pp. 203-205). The resistance pattern occurs when the scientist refuses to accept his own or someone else's discovery because of theoretical or methodological preconceptions, the force of superior authority or prestige, or the prejudices of particular schools of thought (Barber & Hirsch 1962, chapter 32).

Of course, there has always been some resistance to scientific discoveries on the part of cultural and social institutions outside science. Religion, political and economic interests—indeed, the whole range of social factors that interact with science—have under some conditions opposed one or another scientific idea or requirement, just as they have also supported others. An important task for the sociology of science is to analyze specific sources of acceptance and resistance for specific types of scientific ideas and needs.

### Science as a social problem

Like the relations among any of the parts of society, the relations between science and various other parts of society are often inharmonious, as the participants in these relations see them, or

dysfunctional, as the objective observer sees them. Thus, sociology can also study science as what it calls, in general, "a social problem." There are two aspects of science as a social problem. In one aspect, science is an activity whose participants feel, in some important measure, angry or hostile to society because their values and needs are not being properly met or respected. For example, scientists resent what they define as unnecessary controls on their work: they complain about insufficient financial support for their preferred types of research; they deplore politically imposed secrecy; and they try to throw off all nationalistic and other parochial limitations on their research and communication activities. However much these frustrations may be necessary in the light of other social needs than those of science, to scientists they are social problems calling for protest and reform. A variety of organized and unorganized means of protest now exist among scientists to cope with these problems.

In another aspect, science is a social problem to various social groups who consider some of its consequences harmful to themselves and who therefore want to restrict or even eliminate science. Over both the short run and the long, many of the consequences of science are in fact harmful to various social groups. However unintended in most instances, these harmful consequences raise the issue of the social responsibilities of scientists, who are in part their source. It has come to be seen that responsibility for the harmful social consequences of science belongs not directly with scientists but with the several established social and political processes for handling social problems. It has also come to be seen that scientists can play a variety of useful roles in these social and political processes. If scientists have no absolute responsibility for the troubles they help to bring, neither are they absolved from all concern. A number of social and political arrangements are now being worked out to permit scientists to act as expert advisers on the technical aspects of the social problems they have helped to bring into being or that they can foresee arising out of their activities.

BERNARD BARBER

[See also DIFFUSION; INNOVATION; KNOWLEDGE, SOCIOLOGY OF; MEDICAL CARE, article on ETHNOMEDICINE; TECHNOLOGY; and the biographies of OGBURN; WEBER, MAX; ZNANIECKI.]

### BIBLIOGRAPHY

- BARBER, BERNARD 1952 *Science and the Social Order*. Glencoe, Ill.: Free Press. → A paperback edition was published in 1962 by Collier.



- BARBER, BERNARD 1956 *Sociology of Science: A Trend Report and Bibliography*. *Current Sociology* 5, no. 2.
- BARBER, BERNARD; and HIRSCH, WALTER (editors) 1962 *The Sociology of Science*. New York: Free Press. → Contains 38 selections and a bibliography.
- BEARDSLEE, DAVID C.; and O'DOWD, DONALD D. 1961 The College-student Image of the Scientist. *Science* 133:997-1001.
- BEN-DAVID, JOSEPH 1960 Scientific Productivity and Academic Organization in Nineteenth-century Medicine. *American Sociological Review* 25:828-843.
- DEDIJER, STEVAN 1961 Why Did Daedalus Leave? *Science* 133:2047-2052.
- DUPREE, A. HUNTER 1959 *Asa Gray: 1810-1888*. Cambridge, Mass.: Harvard Univ. Press.
- GILLISPIE, CHARLES C. 1951 *Genesis and Geology: A Study in the Relations of Scientific Thought, Natural Theology, and Social Opinion in Great Britain, 1790-1850*. Harvard Historical Studies, Vol. 58. Cambridge, Mass.: Harvard Univ. Press. → A paperback edition was published in 1959 by Harper.
- GILLISPIE, CHARLES C. 1960 *The Edge of Objectivity: An Essay in the History of Scientific Ideas*. Princeton Univ. Press.
- GLASER, BARNEY G. 1964 *Organizational Scientists: Their Professional Careers*. Indianapolis, Ind.: Bobbs-Merrill.
- HAGSTROM, WARREN O. 1965 *The Scientific Community*. New York: Basic Books.
- KUHN, THOMAS S. 1962 *The Structure of Scientific Revolutions*. Univ. of Chicago Press.
- LURIE, EDWARD 1960 *Louis Agassiz: A Life in Science*. Univ. of Chicago Press.
- MALINOWSKI, BRONISLAW (1916-1941) 1948 *Magic, Science and Religion, and Other Essays*. Glencoe, Ill.: Free Press. → A paperback edition was published in 1954 by Doubleday.
- MEAD, MARGARET; and MÉTRAUX, RHODA 1957-1958 Image of the Scientist Among High-school Students. *Science* 126:384-390, 1200, 127:349-351.
- MENZEL, HERBERT (1959) 1962 Planned and Unplanned Scientific Communication. Pages 417-441 in Bernard Barber and Walter Hirsch (editors), *The Sociology of Science*. New York: Free Press.
- MERTON, ROBERT K. (1936) 1962 Puritanism, Pietism and Science. Pages 33-66 in Bernard Barber and Walter Hirsch (editors), *The Sociology of Science*. New York: Free Press.
- MERTON, ROBERT K. (1939) 1957 Science and Economy of 17th Century England. Pages 607-627 in Robert K. Merton, *Social Theory and Social Structure*. Glencoe, Ill.: Free Press. → First published in Volume 3 of *Science and Society*.
- MERTON, ROBERT K. 1957 Priorities in Scientific Discovery: A Chapter in the Sociology of Science. *American Sociological Review* 22:635-659.
- MERTON, ROBERT K. 1961 Singletons and Multiples in Scientific Discovery. *American Philosophical Society, Proceedings* 105:470-486.
- OGBURN, WILLIAM F. (1922) 1950 *Social Change, With Respect to Culture and Original Nature*. New edition with supplementary chapter. New York: Viking.
- PARSONS, TALCOTT 1951 *The Social System*. Glencoe, Ill.: Free Press.
- PRICE, DEREK J. DE Solla 1963 *Little Science, Big Science*. New York: Columbia Univ. Press.

- ZNANIECKI, FLORIAN 1940 *The Social Role of the Man of Knowledge*. New York: Columbia Univ. Press.
- ZUCKERMAN, HARRIET A. 1965 Nobel Laureates in the United States: A Sociological Study of Scientific Collaboration. Ph.D. dissertation, Columbia Univ.

## IV

## SCIENCE-GOVERNMENT RELATIONS

The spectacular achievements of scientists during World War II not only made science obviously more important as an instrument of national policy but also gave it a role of greater independence and initiative in the economic and political systems of the world.

At least since the Enlightenment, it had generally been assumed that the advancement of science would contribute to social and political progress, but the specific role of science, and of scientists, in economic affairs had generally been thought of as subordinate to that of the entrepreneur or manager and, in matters of government policy, to that of the politician and bureaucrat. Especially in the English-speaking countries, the way of thinking initiated by Bacon, and re-emphasized by Franklin and Jefferson, assumed that scientific progress would be maintained by the patient accumulation of facts required for the solution of practical problems, and that research should therefore be supported as incidental to particular industrial enterprises or governmental programs or as a by-product of the system of higher education.

Among leading scientists, this assumption had come to seem less obviously true in the late nineteenth or early twentieth century, in view of the superior accomplishments in fundamental science of various countries in western Europe. The success of Napoleonic France in creating, under governmental auspices, strong institutions for the support of basic science as well as technology, and the state-encouraged systematic advancement, late in the nineteenth century, of research in the universities and industry of Germany, made British and American scientists begin to doubt that the inductive and applied approach in the laboratory and the complete independence of science from political issues and governmental support were still the best means of advancing science.

These doubts were confirmed by the institutional innovations of World War II, which, particularly in the United States, put science in a quite different relation to government. In its military application, it was no longer merely an instrument for the improvement of weapons at the request of military planners: it was a source of independent initiative in the invention of entirely new systems of weapons, to which war planners were forced to adapt

their strategies and diplomats their systems of international relations. Similarly, industrial research or development was no longer to be undertaken only at the initiative of the individual inventor or supported only on the basis of the calculation of the individual corporation; government planners began to look on it as a possible instrument by which government might foster the growth of the national economy and the solution of major social problems, and local communities became eager to foster it as an incentive to the development of their prosperity.

Even more important, political leaders no longer assumed that basic research would be adequately supported as a by-product of the system of industrial enterprise or private education; in the United States, where private universities were most jealous of their independence from political control, the federal government was called on to support, on an unprecedented scale, the basic research carried on in universities and in industrial laboratories. At the same time, scientists and scientific societies began to think of their role in society, not as one of detachment from governmental affairs and not as one of subordination to industrial managers and bureaucrats, but as one of responsibility for the nature of policies and of the institutional system in which science would be fostered and protected.

As a result, the relationship of science to government took on a new degree of significance and interest in the eyes of social scientists.

Since the fostering of science was to be a deliberate policy of government, the sociologists and psychologists were concerned with the identification of the types of personality and the nature of the educational and social environment that would develop creativity and encourage innovation (Research Conference . . . 1963; Roe 1953). In practical terms, they had to deal with the new systems of communication and computation, the management of laboratories, the relationship of scientists to administrators, and the status of professional and scientific groups in the social and political system (Hill 1964).

Since science was considered a potent force for economic growth, economists were led to study the criteria for the allocation of resources for the support of research and development (Nelson 1959). The relation of science and technology to the business cycle, the processes of innovation, and the relationship of business to government in the advancement of technology posed problems quite different from those of either classical or Marxian economics.

Students of politics and administration were

similarly required to consider how far the new techniques of operations research and other methods developed by scientists could go in determining political issues, and what effect they would have on the traditional values and on legal and political processes. These theoretical questions had their practical parallels within government: the relationship of scientific to administrative personnel within the civil service, the status of scientists as advisers to political executives or as participants in political action, and the extent to which representative legislative bodies should organize themselves to deal directly with scientists and with issues concerning the support and application of science (Gilpin & Wright 1964; McCamy 1960).

### Major areas of concern

Problems like those mentioned above had their greatest impact, of course, in the major nations, notably the United States, the United Kingdom, the nations of western Europe, and the Soviet Union, but during the 1950s they also began to be a concern of those interested in the economic and political development of other countries. Let us note three of the major fields in which these new problems arose and then consider their effect on scientific institutions, on governmental institutions, and on the nature of their relationship.

**Science and military strength.** The success of scientists in creating weapons that were not merely improvements on existing weapons systems but entirely beyond the imagination of military planners gave scientists quite a different status in government. Military strength could no longer be measured mainly in terms of mass armies and a system of industrial mobilization for the manufacture of munitions. It depended on the rate of scientific advance and the innovating ability of scientific institutions.

Accordingly, a substantial portion of the budgets of major countries came to be devoted to military research and development programs, and scientists and scientific institutions came to have a large share in the control of such programs.

As a result, the expenditures for research and development by government increased sharply, both in actual amounts and in proportion to other sources of support. In the United States, for example, total expenditure on research and development in 1940 (public and private) was well under \$1 billion, the government share of the total was certainly less than one-fifth, and the military services provided much less than the Department of Agriculture. In 1964 the federal government provided about \$15 billion, roughly two-thirds of



the national total, but spent four-fifths of its funds through private corporations and universities. About \$14 billion of the government total was contributed by the Department of Defense, the Atomic Energy Commission, and the National Aeronautics and Space Administration, the three agencies whose motive for expenditure was most clearly the strengthening of national power and international prestige (U.S. National Science Foundation, *Federal Funds* . . .).

The newest and most powerful weapons systems have been developed, not in the traditional military laboratories, but in technological institutions directed by civilians. In the United States during World War II, the Office of Scientific Research and Development, a civilian agency within the Executive Office of the President, set up an extensive network of contractual relations with universities and industrial corporations, and the Manhattan District of the Army Engineers, which finally developed the atomic bomb, followed the same general pattern (Hewlett & Anderson 1962). As a result, the postwar structure of weapons research and development was one involving heavy reliance on private institutions by the military services, the Atomic Energy Commission, and (later) the National Aeronautics and Space Administration. In the planning of these programs, considerable influence was exercised by advisory committees representing the private scientific institutions. Their influence was greater in the decisions relating to the support of basic research, but it carried considerable weight even in the development of weapons systems (Price 1954).

Scientists in wartime proved to be useful not merely in developing particular weapons but also in devising more effective systems and procedures of tactics and strategy. Operations research applied rigorous and quantitative methods to the solution of problems faced by the military commander, the industrial manager, or the political executive. Such work was organized either within the regular framework of the military services or under the sponsorship of independent research corporations. The flexibility of the contracting method in the United States was shown when the government proceeded to sponsor the creation of private corporations, supported by government funds, for the purpose of assisting governmental authorities in making major decisions on government administration or policy (U.S. Bureau of the Budget 1962). Of these new institutions the RAND Corporation, set up to serve the Air Force, was perhaps the best known, and the techniques that it developed, drawing on the skills of social as well as physical sci-

tists, became important tools for the civilian executives in the Defense Department who sought, in the 1960s, to exercise more effective control over military affairs. [See MILITARY POLICY.]

**Science and economic growth.** Soon after World War II, even those countries committed to the preservation of private economic enterprise had also committed themselves to a policy of maintaining a high degree of employment and fostering economic growth and prosperity. To these countries and others, it was obvious that some of the most dynamic economic growth was taking place in branches of industry which had been stimulated by the advance of technology for military purposes—notably, electronics, aircraft, plastics, and computers.

Economists accordingly took note of the extent to which scientific research and technological development, as well as capital investment, contributed to economic growth. Accordingly, one of the principal motives which led governments to support scientific research was the belief that research and technical education were major factors in the increase in productivity (Denison 1962). Even though economists may not agree on the precise percentage of economic growth which may be attributed to technological advance and the improvement in technological and managerial education, the belief in their efficacy has helped to persuade the United States and most nations of western Europe to increase their research and development expenditures in recent years at a much more rapid rate than their gross national product (Organization for Economic Cooperation and Development 1963, p. 22). In the United States the state governments and local business organizations have been persuaded that federal research subsidies have much to do with the location of new industries in their respective areas, and accordingly the expansion of universities and research laboratories has taken on an importance within the American political system much like that of the traditional forms of public works (such as the improvement of harbors or the building of dams).

At the same time, economists have become more skeptical about the extent to which research undertaken for military purposes will have useful by-products for the civilian economy. In western Europe, where various countries have undertaken to plan the development of the economy, including those sectors carried on by private industry, efforts have been made to develop scientific research as a means of economic growth (Parliamentary and Scientific Conference . . . 1965). The National Economic Development Council of Great Britain and

the Commissariat au Plan of France, for example, have set up special relationships with their respective national organizations for the planning of scientific research expenditures.

In the United States, on the other hand, it proved more difficult to build up federal government research and development programs specifically for the purpose of assisting in the development of private industry. The opposition of private business interests restrained the expansion of certain programs of scientific research and technological assistance in the Department of Commerce, for example, even while federal research support was being increased in other fields relating to the private economy, notably in agriculture and natural resources.

The apparent influence of science and technology in advancing economic growth led naturally to the hope that science could play an important role in the advancement of the underdeveloped countries. This hope found expression in the programs of the several specialized agencies of the United Nations, notably UNESCO, the Food and Agriculture Organization, and the World Health Organization. It became an important element also in the foreign aid and technical assistance programs of various individual nations—notably the United States, which was able to draw on its century of experience in agricultural education and research. [See ECONOMIC GROWTH.]

**The health sciences.** Some of the most striking practical accomplishments of the sciences have been in the fields of medicine and public health, and the medical profession and hospitals accordingly have been involved in the research programs supported by government to a rapidly increasing degree.

In the United States the National Institutes of Health (NIH) of the U.S. Public Health Service became during the 1950s the principal supporter of biological and medical sciences, not merely in the United States but throughout the world. American universities were receiving grants for basic research in even larger amounts from the NIH than from the military services or the National Aeronautics and Space Administration, and NIH grants throughout Europe had begun to add significantly to the research funds available to university scientists and to encourage the development of new patterns of research organization.

The official health programs supported by grants from the U.S. government and from the World Health Organization continued and extended the earlier international programs for the eradication of disease of such charitable organizations as the

Rockefeller Foundation and developed them further with the help of techniques developed or advanced by wartime research programs. The success of such efforts in reducing the death rate intensified concern with the problem of a population increasing at a rate which seemed to threaten the food supply and the economic progress of the various underdeveloped regions, especially the newly independent countries of Asia and Africa. [See PUBLIC HEALTH.]

### Institutions of basic science

Universities, which in their organization still retain traces of their medieval ecclesiastical origin, are in most countries the principal institutions for the conduct of basic research. In the major countries of continental Europe, they generally came under the patronage of the state, and their programs of scientific and technological research were seen as serving national purposes. Particularly in France and Germany, strong centers of fundamental research were thus developed with government support at a time when the ancient British universities were still dominated by classical studies and were generally uninterested in the advancement of science, and American institutions of higher learning were either denominational colleges or state institutions preoccupied with comparatively low levels of vocational and technical education (Ashby 1958; Cohen 1963).

The twentieth-century progress of science in the United Kingdom came in the basic sciences and mathematics at Cambridge (and later Oxford) and even more notably in some of the newer universities that even in the nineteenth century had been concerned with technological training and the needs of industry. A major share of the cost of these universities was taken over by the national treasury but was channeled through the University Grants Committee in such a way as to insulate the universities from political control.

In the United States, basic research before World War II was carried on mainly by the major universities, private or state, which received extensive assistance from the philanthropic foundations in the form of grants for special research projects. The leadership of the scientific societies and some of the philanthropic foundations had, by World War II, been persuaded that the nation should greatly increase its emphasis on fundamental research (U.S. National Science Foundation 1960). The patterns that had been developed by the grant-making programs of the private foundations were influential in determining the system of government support for such research in private and state



universities after 1945. That system supplied funds from several executive departments and agencies to universities and private corporations, under a system in which panels and committees of advisers from the private institutions were given great influence in the allocation of grants for specific projects. At the same time, the President's Science Advisory Committee, its members drawn mainly from private institutions, put great emphasis on a policy of keeping basic research and the teaching of graduate students closely connected within the universities, a policy that in addition to its pedagogical advantages had the merit of strengthening the autonomy of the academic and scientific community (U.S. President's Science Advisory Committee 1960).

The Soviet Union, on the other hand, chose to divide the functions of higher education from those of advanced research and generally to assign them to separate institutions. Research was directed through a system headed by the Committee on the Coordination of Research and Development, the Academy of Sciences, and corresponding academies for the several constituent republics (DeWitt 1961, pp. 434-435). In theory, the entire program is under the guidance of the principles of dialectical materialism; in practice, this guidance seems to affect the conduct of research least in the physical sciences, more in the biological sciences, and still more in the social sciences and humanities.

### Scientists in administration and politics

Basic as well as applied research is conducted not merely in universities but also in government laboratories set up to further the purposes of particular departments or ministries. The way in which their work, as well as the work of scientists in industry and universities, will be related to government policy and government programs depends to a great extent on the status, within the political system and the administrative service, of men trained in the sciences and related professions.

In the Soviet Union, the break with traditional systems of education and the great emphasis on scientific and technological training cleared the way for the advancement of scientists and engineers to positions in the higher ranks of the bureaucracy, and their advancement was further encouraged by the absence of separate cadres of private business managers and of institutions for their education. By contrast, the career administrative systems of western Europe and the United Kingdom have been comparatively resistant to the promotion of scientists and engineers to high administrative positions. In Germany, for example, a

legal education remains the normal basis for administrative advancement, and in Great Britain the highest class of the civil service is still dominated by permanent officials whose education was in history, the classics, and related fields of knowledge.

The personnel system of the United States, which lacks an elite corps and encourages exchanges between the civil service and private life at all levels in the hierarchy, has made it comparatively easy for scientists and men with technological education to rise to high administrative positions. The continuous interchange between governmental and private careers makes the government service a much less distinct part of society, and the distinction between governmental and private careers is further confused by the recent growth of the practice of contracting with private corporations and universities for the performance of government functions, especially research and development programs. As a result, the leading scientists, based in private institutions, who are called on to serve in a part-time or advisory capacity within the structure of government acquire an unprecedented degree of influence.

In the United States this advisory function takes the institutional form of a hierarchy of advisory committees, some of a continuing nature and some *ad hoc*, set up in the principal civilian agencies and military services, and in the White House itself. The most difficult administrative problems relating to the use of such committees have to do with their relationship to the main line of administrative authority, including the full-time staff. The first tendency after World War II was to look to the part-time advisers alone for the principal policy and administrative decisions, notably in the major advisory committee attached to the Executive Office of the President and the Office of the Secretary of Defense. But with the abolition of the Research and Development Board of the Department of Defense and the eventual substitution of a much more powerful individual officer, the director of defense research and engineering, and with the transformation of the President's Science Advisory Committee by the provision of a full-time chairman supported by a full-time staff, a more effective relationship was devised between the part-time adviser and the full-time official, and more sophisticated concepts were worked out with respect to the particular ways in which part-time advisers could most usefully be employed (Gilpin & Wright 1964).

The entire system is under the powerful influence, and often the detailed control, of congressional committees that have been making direct

use of independent scientific advisers, especially through the mechanism of the National Academy of Sciences. In countries with parliamentary forms of government the general response to a similar need was the creation of interministerial committees to deal with government-wide scientific problems, often drawing on the advice of committees including private scholars. Thus in the Federal Republic of Germany an interministerial committee exists side by side with an advisory council on science and the humanities and a committee on cultural policy and publicity of the parliament. Similarly, in France an interministerial committee reporting to the premier is advised by a consultative committee of scientists and economists, both committees being served by a single secretariat (the *Délégation Générale à la Recherche Scientifique et Technique*). The *Délégation Générale* works closely with the national planning agency (Parliamentary and Scientific Conference . . . 1965).

In the United Kingdom the Advisory Council on Scientific Policy was set up to advise the lord president of the council, who was responsible for the general program of civilian science in Great Britain. The lord president's responsibilities were later transferred to a newly created minister for science, whose functions were divided in 1964 by the newly elected Labour government between a minister for science and education and a minister for technology. Unlike the United States or the major countries of western Europe, the United Kingdom has no standing parliamentary committees dealing with particular scientific and technological programs, but it has the informal Parliamentary and Scientific Committee, including members of both houses of Parliament and representatives of more than one hundred scientific and technical organizations, to further the processes of consultation and the exchange of information between the worlds of politics and science.

The interest of national governments in science policy has been furthered by studies conducted by various international organizations. The Organization for Economic Cooperation and Development, the United Nations Educational, Scientific and Cultural Organization, and the North Atlantic Treaty Organization have all urged more comprehensive and consistent policies for the support and the coordination of scientific programs, and for more responsible forms of organization to deal with science at high policy levels.

Similar international discussions led to the formation of more effective international associations for the promotion of the several sciences, most

of them being grouped under the International Council of Scientific Unions with the support of UNESCO, and to the formation of international organizations for the conduct of scientific programs. Some of these latter organizations were set up for the operation of specific temporary programs, such as the International Geophysical Year, and others provided regional organizations for tasks too expensive or complex for single nations to undertake (for example, the European Organization for Nuclear Research, the European Atomic Energy Community, and the European Space Research Organization).

The conviction that scientists have a moral and political responsibility for the consequences of their research led many of them to abandon their old posture of political indifference or neutrality and to organize to take part in public affairs not merely as subordinate advisers but as active participants. This conviction led to the formation of many organizations for public education and the expression of policy opinions, their status and methods of operation depending greatly on the degree of freedom permitted to voluntary associations by the philosophy and traditions of the various countries. In the United Kingdom and western Europe, the leadership of such organizations tended before World War II toward a socialist or Marxist political philosophy. Since 1946 this tendency has been counterbalanced by theoretical arguments and organizing efforts in support of scientific freedom and against political control of research or economic enterprise. A good example is the program of the Congress for Cultural Freedom. In the United States, the political activity of scientists has been less influenced by ideology and less committed to particular political parties; it has concentrated more on specific policy issues and on the fostering of interest in public affairs within the scientific community (Price 1965).

DON K. PRICE

[See also NATIONAL SECURITY; POLICY SCIENCES; TECHNOLOGY, article on TECHNOLOGY AND INTERNATIONAL RELATIONS.]

#### BIBLIOGRAPHY

*Some of the principal sources in the field are government reports and documents. International Science Report deals with the organization of science in various countries (the first two reports dealt with India and Germany.) The Organization for Economic Cooperation and Development has published a series of Country Reviews on the policies for science and education, and a series of Country Reports on the organization of scientific research of various European countries. The most useful bibliographies are U.S. National Science Foundation, Office of Special Studies 1959 and Current Projects in Economic*



and Social Implications of Science and Technology, a survey issued annually since 1957 by the foundation. U.S. data on the financing of research appear in the annual series of the U.S. National Science Foundation, Federal Funds for Research, Development, and Other Scientific Activities. In the United States a valuable additional source of data is the hearings and reports of various congressional committees, notably the Appropriations Committee of each house, the House Committee on Science and Astronautics, the Senate Committee on Aeronautical and Space Sciences, the committees of the two houses on government operations, and the Joint Committee on Atomic Energy. Discussions of continuing problems may be found in the periodicals *Bulletin of the Atomic Scientists*, *Minerva*, and *Science*.

ASHBY, ERIC 1958 *Technology and the Academics: An Essay on Universities and the Scientific Revolution*. London and New York: Macmillan.

*Bulletin of the Atomic Scientists*. → Published since 1945.

COHEN, I. BERNARD 1963 *Science in America: The Nineteenth Century*. Pages 167-189 in Arthur M. Schlesinger, Jr. and Morton White (editors), *Paths of American Thought*. Boston: Houghton Mifflin.

DENISON, EDWARD F. 1962 *The Sources of Economic Growth in the United States and the Alternatives Before Us*. New York: Committee for Economic Development.

DEWITT, NICHOLAS 1961 *Education and Professional Employment in the U.S.S.R.* Washington: National Science Foundation.

DUPREE, A. HUNTER 1957 *Science in the Federal Government: A History of Policies and Activities to 1940*. Cambridge, Mass.: Belknap.

GILPIN, ROBERT; and WRIGHT, CHRISTOPHER (editors) 1964 *Scientists and National Policy-making*. New York: Columbia Univ. Press.

GREAT BRITAIN, COMMITTEE OF ENQUIRY INTO THE ORGANISATION OF CIVIL SCIENCE 1963 [Trend Report]. Papers by Command, Cmnd. 2171. London: H.M. Stationery Office.

HEWLETT, RICHARD G.; and ANDERSON, OSCAR E. JR. 1962 *A History of the United States Atomic Energy Commission*. Volume 1: The New World: 1939-1946. University Park: Pennsylvania State Univ. Press.

HILL, KARL (editor) 1964 *The Management of Scientists*. Boston: Beacon.

HITCH, CHARLES J. 1958 *Character of Research and Development in a Competitive Economy*. Pages 129-139 in Conference on Research and Development and Its Impact on the Economy, Washington, D.C., 1958. *Proceedings*. Washington: National Science Foundation.

HITCH, CHARLES J.; and MCKEAN, R. N. 1960 *The Economics of Defense in the Nuclear Age*. Cambridge, Mass.: Harvard Univ. Press.

HOGG, QUINTIN M. 1963 *Science and Politics*. London: Faber. → The author is Viscount Hailsham.

*International Science Report*. → Published since 1962 by the National Science Foundation, Office of International Science Activities.

KIDD, CHARLES V. 1959 *American Universities and Federal Research*. Cambridge, Mass.: Belknap.

KUHN, THOMAS S. (1962) 1963 *The Structure of Scientific Revolutions*. Univ. of Chicago Press.

MCCAMY, JAMES L. 1960 *Science and Public Administration*. University: Univ. of Alabama Press.

MERTON, ROBERT K. (1936) 1962 *Puritanism, Pietism and Science*. Pages 33-66 in Bernard Barber and

Walter Hirsch (editors), *The Sociology of Science*. New York: Free Press.

*Minerva* (London). → Published since 1962.

NATIONAL RESEARCH COUNCIL, COMMITTEE ON SCIENCE AND PUBLIC POLICY 1964 *Federal Support of Basic Research in Institutions of Higher Learning*. Publication No. 1185. Washington: National Academy of Sciences-National Research Council.

NELSON, RICHARD R. 1959 *The Economics of Invention: A Survey of the Literature*. *Journal of Business* 32: 101-127.

ORGANIZATION FOR ECONOMIC COOPERATION AND DEVELOPMENT 1963 *Science, Economic Growth and Government Policy*. Paris: The Organization.

ORGANIZATION FOR ECONOMIC COOPERATION AND DEVELOPMENT, ADVISORY GROUP ON SCIENCE POLICY 1963 *Science and the Policies of Governments: The Implications of Science and Technology for National and International Affairs*. Paris: The Organization.

ORGANIZATION FOR ECONOMIC COOPERATION AND DEVELOPMENT *Country Reports on the Organisation of Scientific Research*. → Published since 1963.

ORGANIZATION FOR ECONOMIC COOPERATION AND DEVELOPMENT *Country Reviews*.

PARLIAMENTARY AND SCIENTIFIC CONFERENCE, SECOND, VIENNA, 1964 1965 *Science and Parliament*. Paris: Council of Europe and Organization for Economic Cooperation and Development.

PECK, MERTON J.; and SCHERER, FREDERICK M. 1962 *The Weapons Acquisition Process: An Economic Analysis*. Boston: Harvard Univ., Graduate School of Business Administration, Division of Research.

PRICE, DON K. 1954 *Government and Science: Their Dynamic Relation in American Democracy*. New York Univ. Press. → A paperback edition was published in 1962 by Oxford University Press.

PRICE, DON K. 1965 *The Scientific Estate*. Cambridge, Mass.: Belknap.

RESEARCH CONFERENCE ON THE IDENTIFICATION OF CREATIVE SCIENTIFIC TALENT 1963 *Scientific Creativity: Its Recognition and Development*. Edited by Calvin W. Taylor and Frank Barron. New York: Wiley.

ROE, ANNE 1953 *The Making of a Scientist*. New York: Dodd

*Science*. → Published since 1883 by the American Association for the Advancement of Science.

SNOW, CHARLES P. 1961 *Science and Government*. Cambridge, Mass.: Harvard Univ. Press.

UNIVERSITIES-NATIONAL BUREAU COMMITTEE FOR ECONOMIC RESEARCH 1962 *The Rate and Direction of Inventive Activity. Economic and Social Factors*. National Bureau of Economic Research, Special Conference Series, No. 13. Princeton Univ. Press.

U.S. BUREAU OF THE BUDGET 1962 *Report to the President on Government Contracting for Research and Development*. Washington: Government Printing Office.

U.S. NATIONAL SCIENCE FOUNDATION 1960 *A Study of Scientific and Technical Manpower; A Program of Collection, Tabulation, and Analysis of Data of the National Science Foundation: A Report . . .* U.S. House of Representatives, 86th Congress, 2d Session. Washington: Government Printing Office.

U.S. NATIONAL SCIENCE FOUNDATION *Federal Funds for Research, Development, and Other Scientific Activities*. → Published since 1950/1951. Volumes 1-12 were published as *Federal Funds for Science*.

- U.S. NATIONAL SCIENCE FOUNDATION, OFFICE OF SCIENCE INFORMATION SERVICE *Current Projects in Economic and Social Implications of Science and Technology*. → Published since 1957.
- U.S. NATIONAL SCIENCE FOUNDATION, OFFICE OF SPECIAL STUDIES 1959 *Bibliography on the Economic and Social Implications of Research and Development*. Washington: The Foundation.
- U.S. OFFICE OF SCIENTIFIC RESEARCH AND DEVELOPMENT 1945 *Science: The Endless Frontier*. Washington: Government Printing Office.
- U.S. PRESIDENT'S SCIENCE ADVISORY COMMITTEE 1960 *Scientific Progress, the Universities and the Federal Government: Statement*. Washington: Government Printing Office.

## V

## SCIENTISTS

The word "scientist" was introduced into the English language around 1840 to distinguish those who seek empirical regularities in nature from philosophers, scholars, and intellectuals in a more general sense (Ross 1962). Mathematicians and logicians are usually regarded as scientists, although mathematics ceased being regarded as an empirical science by 1890–1910, and today the rubric also covers specialists in the social sciences almost without qualification. Other European languages have no terms strictly equivalent to "scientist." The French *savant*, Italian *scienziato*, German *Wissenschaftler*, and Russian *uchëny* also refer to philosophers, historians, and other systematic scholars. The absence of a verbal distinction is sometimes reflected in the organization of science and scholarship; for example, philosophers and historians are included in the Soviet Academy of Sciences.

Persons in a wide variety of statuses and roles are described as scientists or identify themselves as scientists in English-speaking societies today. In a narrow sense, a scientist is a man of scientific knowledge—one who adds to what is known in the sciences by writing articles or books. This is perhaps the only sense in which the word should be used without qualification. However, "scientists" also engage in applied research, attempting to make discoveries which will lead to new industrial, medical, and agricultural products or processes; in industrial development, applying scientific knowledge to specific problems of innovation in production; and in the routine testing and analyzing of commodities and processes. Such persons are called applied scientists and are not easily distinguished from engineers and technicians. Persons called scientists may also engage in teaching science in institutions of higher education, in writing accounts of science for laymen, and in administration. Finally, a scientist may be defined as a person

who has received a college degree in a scientific field.

**The population of scientists.** Estimates of the number of scientists vary according to the definition of the term and are, in any case, difficult to make. In the United States around 1962, there were more than one million persons with scientific or technical degrees, more than 275,000 members of professional scientific societies, more than 118,000 persons listed in *American Men of Science*, and roughly 100,000 engaged in basic or applied scientific research. The number of persons making substantial contributions to knowledge is, of course, much less than 100,000. The United States leads the world in the number of scientists. Between one-fourth and one-third of all scientists may be Americans, and most of the remainder are in the other industrialized nations. (For example, seven leading industrial nations accounted for more than 80 per cent of the articles abstracted in *Chemical Abstracts* in 1960. There is naturally an even greater concentration of scientific languages: English, German, Russian, and French accounted for more than 90 per cent of the more than ten thousand articles abstracted in *Index Chemicus* for 1963.) Many scientists move to the United States from other nations, industrialized as well as underdeveloped, to obtain better conditions for performing research; this concentration has caused concern among leaders in many nations.

The numbers cited above are changing rapidly. The growth in the number of scientists since the later Middle Ages has been exponential; thus, the scientists living in 1960 probably constituted more than 90 per cent of all those who ever lived (Price 1963). The rate of growth around 1960 would double the number of research scientists in roughly fifteen years. This rate of growth, however, must decline within the next generation in developed nations.

In the United States in 1952, about half of the science doctorates were received in the physical sciences and mathematics, about 22 per cent in the biological sciences, and about 28 per cent in the social sciences, including psychology. The relative proportions in these major subareas of science have not changed much in the last half century. Social scientists constitute a much larger proportion of scientists in the United States than they do in most other nations.

**Evolution of scientific roles.** The concept of the scientist developed only with the professionalization of science in the first half of the nineteenth century; previously, scientific activities were a subsidiary aspect of other social roles. Florian Zna-



niecki (1940) has written a concise sociological account of the historical differentiation of scientific roles. In most preindustrial societies, science has been the activity of practical actors. In ancient and medieval Europe, scientists were academic and often religious teachers. Beginning about the time of Galileo's move in 1611 from the University of Padua to the court of Cosimo de' Medici in Florence, scientists left the universities, and during the seventeenth and eighteenth centuries the amateur scientist flourished: either the gentleman scientist, like Robert Boyle, or the middle-class amateur, like Joseph Priestley. Scientists again became established in universities after the French Revolution. The twentieth century has seen the rapid growth of research establishments in business firms and government agencies. Of Americans receiving the doctorate in 1958, 56 per cent of the physical scientists, 31 per cent of the biological scientists, and 19 per cent of the social scientists took employment in industry or government. In the United States, most basic research is conducted in academic settings, but most of it is conducted in nonacademic research institutes in the Soviet Union and in some western European nations.

### Scientific careers

Roughly 90 per cent of American physical scientists, and slightly smaller proportions of biological and social scientists, are men. Like other professionals, they are recruited disproportionately from upper-middle-class families; but, unlike others, they are much less likely to be from Roman Catholic families and more likely to be from Protestant or Jewish families than would be expected by chance. During the interwar years, a small number of small liberal arts colleges contributed a disproportionately large share of those who went on to become scientists. Much of this resulted from selective recruitment into these colleges, but some probably resulted from their distinctive scholarly ethos. The growth of scientific education has been accompanied by a growth in the proportion of scientists produced by large public and private undergraduate institutions.

**Education.** Scientific education was revolutionized in nineteenth-century Germany by the dissertation: the requirement that the student conduct original research before receiving a degree. This innovation was soon copied in other nations. Today scientific education is prolonged and highly specialized, especially in its later stages. The dissertation research is usually conducted in a quasi-apprenticeship relation with a university professor, and students often make significant contributions to the research of their professors. Almost all

American graduate students in the sciences receive stipends for this work. The median number of years elapsed between the bachelor's degree and the doctorate for Americans receiving the doctorate in 1957 was six in the physical sciences, seven in the biological sciences, and eight in the social sciences (see Berelson 1960 on this and other aspects of graduate education).

Although many American universities give advanced training in the sciences (110 gave doctorates in chemistry in 1960), students tend to be concentrated in those institutions best known for basic research; the 15 universities with highest prestige accounted for almost half the doctorates in the sciences in 1957. The best graduate schools tend to recruit students from the best undergraduate schools, and their students are much more likely to be employed in leading universities than are graduates of institutions of lesser renown. Among those receiving advanced degrees, those who do best academically and identify most closely with their professors are most likely to go on to do research and teaching in universities; government and industrial laboratories tend to recruit those who do less well and are less committed to a research career.

**Career mobility.** There is now considerable mobility between universities and research establishments in government and industry; for example, more than one-third of American chemists have been employed in at least two of the three locales (Strauss & Rainwater 1962, chapter 6). The barriers to mobility of this sort were much greater before World War II and apparently are still great in most of the countries of western Europe. There is much less mobility between the major fields of science because of the specialized education required to enter any one of them.

Scientists in all types of establishments are likely to combine research with other activities; a large majority of scientists combine research with such other activities as teaching, administration, and technical consultation. Vertical mobility takes different forms in different locales. Advancement in universities does not usually involve a qualitative change in the nature of the work; it still involves teaching and research. In government, and especially in industry, vertical mobility often involves advancement to administrative positions and the abandonment of research; many firms recruit managers for line operations from among their research scientists (Kornhauser 1962, chapter 5).

### Incentives and occupational personality

Scientists in basic research are strongly motivated to solve intellectual problems that they re-

gard as intrinsically important, but the most important *social* incentive for them is their desire to obtain recognition from their colleagues for their research accomplishments. This desire leads the scientist to publish his results, and it influences his decisions in the selection of research problems and methods. Scientists compete strenuously to be the first to publish discoveries, and simultaneous discoveries by two or more scientists occur frequently. When there is some question about which scientist made a discovery first, a priority dispute may arise. The frequency of such disputes and the intense bitterness which often accompanies them are telling evidence of the value scientists place upon the esteem of their colleagues (Merton 1957; 1963); however, priority disputes have become less frequent in the twentieth century. Industrial and governmental research establishments which conduct applied or secret research make it more difficult for their scientific employees to be recognized, and such establishments must offer relatively high salaries to scientists in order to induce them to accept this deficiency.

**Values and personality.** Compared with non-scientists, scientists are more inclined to prize the recognition of their colleagues and their professional autonomy above the rewards of income, organizational power, and community prestige. They are nevertheless among the most prestigious occupational groups. "Scientists" were ranked third in prestige among 90 occupations—just behind U.S. Supreme Court justices and physicians—by a cross section of the U.S. population in 1963; some scientific specialties were, however, ranked a good deal lower. There is some evidence that these distinctive occupational values characterize scientists even as university undergraduates (Rosenberg 1957; Davis 1966).

Many of the distinctive personality characteristics of scientists also appear to be produced more by selection than by university or occupational socialization. Scientists tend to be highly intelligent; possessors of U.S. doctorates in the sciences have a mean IQ of over 130 (from a population with mean 100 and standard deviation 20), which is not much different from the mean IQ of similarly qualified persons in other fields or the mean IQ of medical and law school graduates (Price 1963, chapter 2). Although the evidence is sketchy and incomplete, and although there is much variation among scientists, it also appears that physical and biological scientists tend to be "intensely masculine," to avoid close interpersonal contacts, and to avoid and be disturbed by emotions of hostility and dependency. These characteristics may help generate the scientist's desire for recognition for

his accomplishments from his colleagues. Physical scientists tend to like music and dislike poetry and art. As might be expected, creative scientists are unusually hardworking, to the extent of appearing almost obsessed with their work (see Research Conference . . . 1963 for summaries of studies on the personality characteristics of scientists). Stereotypes of scientists held by high school and university students are in crude accord with these psychological findings (Mead & Métraux 1957–1958; Beardslee & O'Dowd 1961).

It has been argued that the values of the Puritans—rational mastery of one's environment, worldly activity for the glorification of God, and individualism—helped to motivate them to pursue science in seventeenth-century England, and have resulted, even today, in a higher valuation of science among Protestants than among Roman Catholics; this is a matter of dispute, however (Merton 1939; Feuer 1963). In any case, scientists today are among the least religious groups in the American population, with respect to both religious beliefs and religious practices. Those in the behavioral and biological sciences, where religion and science are most likely to come into conflict, are least likely to be religious. Scientists also tend to support parties of the left in national politics. With regard to both religion and politics, there is evidence that the more productive scientists are less conservative than the less productive.

**Scientific productivity.** It is generally difficult to account for variations in scientific productivity in terms of personality or background variables. There is great variation in productivity; the probability that a scientist will produce  $n$  or more papers (if he produces at least one) is roughly proportional to  $1/n$ , and 50 per cent of the papers are written by about 6 per cent of the scientists (Price 1963, table 2, p. 45). The best background predictors of scientific productivity are the times required to get the doctorate and college grades, but these account for little of the variation. Contrary to common belief, productivity does not decline with age until advanced ages, 60 or over, are reached.

Social contexts may have greater effects on productivity, but these effects are not readily distinguished from the selectivity of the contexts themselves. Graduates and faculty of leading universities are more productive than others. Scientists who have adequate facilities and research assistance are more productive than those who do not, and those who maintain continuity in their research topics are more productive than those who do not. However, the kinds of contacts with colleagues and the quality of specialization make a difference. Scientists who perform multiple func-



tions, combining research with teaching or administration, have been shown to be more productive than those who devote full time to research. Scientists who have frequent contacts with colleagues possessing different research interests are more productive than those with few contacts or contacts only with others of very similar interests. Social isolation especially is associated with low productivity for most scientists; this is a major handicap for scientists working in underdeveloped nations.

### The organization of science

Scientists engaged in basic research are necessarily given considerable autonomy. If they are not free to evaluate the truth or falsehood of theories or the adequacy of research findings, science ceases to exist; and scientists also are usually free to select research problems and techniques within broad limits. Such autonomy exists in many universities and in a few governmental and industrial establishments. University departments place some restrictions on this freedom; for example, some research in logic or statistics may be felt to be not "really" mathematics, and scientists in mathematics departments may be discouraged from performing it. This may produce organizational strains, and these strains are sometimes resolved by the differentiation of existing disciplines and departments. The number of recognizably different scientific disciplines probably trebled between 1900 and 1960, and the formation of new disciplines shows no signs of stopping.

The autonomy of basic scientists conflicts with a frequent need to engage in cooperative research. Traditionally this conflict has been resolved by forming temporary teams of freely collaborating scientists or teams of a professor and his students. More recently, in fields like nuclear physics, cooperative research has been formalized; permanent groups of professional scientists are formed in which authority is centralized and a formal division of labor is established. In most fields in universities, however, the traditional forms of teamwork are much more common than such formally organized groups.

**Industry and government.** The importance that scientists place upon purely scientific goals, their desire to be autonomous, and their sensitivity to the responses of their disciplinary colleagues produce strains when they are employed by industrial and governmental agencies to help achieve practical goals. For example, the desire to select their own research problems leads them to resist accepting directions from organization superiors. Also, scientists tend to prefer working in units with their

disciplinary colleagues to working in professionally heterogeneous groups that are organized on the basis of industrial functions; and the desire to inform others of their discoveries conflicts with requirements of industrial and military secrecy. The typical industrial incentives of salary and power are less important for scientists than for other categories of employees, and, if scientists accept these incentives as primary, their commitments to scientific values and their scientific competences may be eroded.

Various ways have evolved for accommodating industrial organization and the typical organization of the scientific community (Kornhauser 1962). Industrial scientists are selected from among those most willing to accept the importance of industrial goals, and occupational socialization furthers this acceptance. In addition, industrial research organizations are often differentiated into fundamental research units and units more directly involved in practical tasks, and scientists are assigned to different units on the basis of interests and skills. Research supervisors are recruited from among superior scientists and tend to use persuasion more than formal direction. "Parallel hierarchies" of advancement may be offered, so that some scientists are promoted to positions giving them greater autonomy in research, while others are promoted to administrative positions. Finally, the patent system and similar procedures make it possible for scientists to publish some research findings while safeguarding the proprietary interests of firms in discoveries. These types of accommodations make possible the incorporation of scientists into industry without the sterilization of science.

**Professional societies.** In addition to their work establishments, scientists as scientists are organized into a wide variety of societies and associations. Most of the vast number of such groups are devoted to a specialized subject matter, and their major activities involve the communication of discoveries through meetings and journals. The oldest societies and those with highest prestige, like the Royal Society of London and the U.S. National Academy of Sciences, are interdisciplinary, and membership in them is by election. Such groups engage in facilitating scientific communication, but they also play important roles in formulating the science policies of national governments. So also do large national groups that are open to almost all scientists, such as the American Association for the Advancement of Science, the British Association for the Advancement of Science, and the Deutscher Verband Technisch-Wissenschaftlicher Vereine.

While there has been a tendency for the scien-

tific "society" to give way to the professional "association," none of these types of organizations is directly involved with questions of wage rates and working conditions. At most, these groups forward the occupational interests of scientists by protecting job titles through state licensing and certification, by elevating educational standards, and by lobbying for increased governmental support. There are a few trade unions for "scientists," but engineers and technologists dominate them and they are relatively ineffective.

### What is a scientist?

The concept of the scientist was originally formulated primarily to distinguish specialists in the natural sciences from philosophers, historians, and other intellectuals; the distinction began to emerge as clear and significant only around 1800. Today the differences between philosophy and an empirical, generalizing, value-free science are clearly recognized. It is known that the pursuit of science requires the full-time efforts of its workers, and provisions for scientific employment are abundant. Perhaps we have succeeded only too well in distinguishing scientists from other intellectuals. Today our major difficulties involve distinguishing scientists from engineers and technicians. Even when a scientist is clearly a man of knowledge, his knowledge is apt to be highly specialized and communicable to few others, even within his own discipline. Are scientists intellectuals? Or are they a culture (or cultures) apart? C. P. Snow's 1959 book, *The Two Cultures and the Scientific Revolution*, stimulated extensive public discussions of this (the essays in Holton 1965 are a good sample of the result).

Perhaps university professors find it too easy to discover a cleavage between scientists and others. Taking a longer view, one might conclude that scientific knowledge and the scientific approach to the world have been dispersed very widely and deeply in the populations of industrialized nations. Most of us have been deeply influenced by scientists in what we believe and how we believe. The willingness of populations to tolerate and support scientists may be some testimonial to this influence.

WARREN O. HAGSTROM

[See also CREATIVITY, article on SOCIAL ASPECTS; ENGINEERING; INTELLECTUALS; KNOWLEDGE, SOCIOLOGY OF; and the biography of ZNANIECKI.]

### BIBLIOGRAPHY

BARBER, BERNARD; and HIRSCH, WALTER (editors) 1962 *The Sociology of Science*. New York: Free Press. →

- The articles by Beardslee & O'Dowd (1961), Mead & Métraux (1957-1958), Merton (1939; 1957) are reprinted in this collection. Includes a bibliography.
- BEARDSLEE, DAVID C.; and O'DOWD, DONALD D. 1961 *The College-student Image of the Scientist*. *Science* 133:997-1001.
- BERELSON, BERNARD 1960 *Graduate Education in the United States*. New York: McGraw-Hill.
- DAVIS, JAMES A. 1966 *Undergraduate Career Decisions*. Chicago: Aldine.
- FEUER, LEWIS S. 1963 *The Scientific Intellectual: The Psychological and Sociological Origins of Modern Science*. New York: Basic Books.
- HAGSTROM, WARREN O. 1965 *The Scientific Community*. New York: Basic Books.
- HARMON, LINSEY R. 1965 *Profiles of Ph.D.'s in the Sciences: Summary Report on Follow-up of Doctorate Cohorts, 1935-1960*. Publication 1293. Washington: National Academy of Sciences-National Research Council.
- HOLTON, GERALD (editor) 1965 *Science and Culture: A Study of Cohesive and Disjunctive Forces*. Boston: Houghton Mifflin. → First published in Volume 94 of *Dædalus*.
- KORNHAUSER, WILLIAM 1962 *Scientists in Industry: Conflict and Accommodation*. Berkeley: Univ. of California Press.
- MEAD, MARGARET; and MÉTRAUX, RHODA 1957-1958 *Image of the Scientist Among High-school Students*. *Science* 126:384-390, 1200; 127:349-351.
- MERTON, ROBERT K. (1939) 1957 *Science and Economy of 17th Century England*. Pages 607-627 in Robert K. Merton, *Social Theory and Social Structure*. Glencoe, Ill.: Free Press. → First published in Volume 3 of *Science and Society*.
- MERTON, ROBERT K. 1957 *Priorities in Scientific Discovery: A Chapter in the Sociology of Science*. *American Sociological Review* 22:635-659.
- MERTON, ROBERT K. 1963 *Resistance to the Systematic Study of Multiple Discoveries in Science*. *Archives européennes de sociologie* 4:237-282.
- PRICE, DEREK J. DE SOLLA 1963 *Little Science, Big Science*. New York: Columbia Univ. Press.
- RESEARCH CONFERENCE ON THE IDENTIFICATION OF CREATIVE SCIENTIFIC TALENT 1963 *Scientific Creativity: Its Recognition and Development*. Edited by Calvin W. Taylor and Frank Barton. New York: Wiley. → Includes a bibliography on pages 391-407.
- ROSENBERG, MORRIS 1957 *Occupations and Values*. Glencoe, Ill.: Free Press.
- ROSS, SYDNEY 1962 "Scientist": The Story of a Word. *Annals of Science* 18:65-85.
- SNOW, C. P. 1959 *The Two Cultures and the Scientific Revolution*. Cambridge Univ. Press.
- SNOW, C. P. 1964 *The Two Cultures: And a Second Look*. 2d ed. Cambridge Univ. Press. → An expanded version of Snow 1959.
- STRAUSS, ANSELM L.; and RAINWATER, LEE 1962 *The Professional Scientist: A Study of American Chemists*. Chicago: Aldine.
- U.S. NATIONAL SCIENCE FOUNDATION 1965 *Scientific and Technical Manpower Resources: Summary Information on Employment Characteristics, Supply and Training*, prepared by Norman Seltzer. NSF 64-28. Washington: Government Printing Office.
- ZNANIECKI, FLORIAN 1940 *The Social Role of the Man of Knowledge*. New York: Columbia Univ. Press.



## VI

## SCIENTIFIC COMMUNICATION

The term "scientific communication" refers to the exchange of information and ideas among scientists in their roles as scientists. Menzel (1958, p. 6) defines it as "the totality of publications, facilities, occasions, institutional arrangements, and customs which affect the direct or indirect transmission of scientific messages among scientists." It is distinguished from everyday communication about physical reality in that it has reference to a particular body of generalized, codified knowledge. Ideally, each and every communication contributes to the corpus of accepted knowledge identified as science. This is accomplished chiefly by extending the boundaries, by modifying previously held hypotheses, and by introducing additional precision, clarification, or verification of existing knowledge.

Although the preferred means and the practices associated with scientific communication have undergone change in the last few centuries, full and open communication of scientific results has always been a foundation stone of modern science. With the establishment of the academies in the seventeenth century, word-of-mouth exchange of information and informal meetings were very quickly supplemented by informal correspondence and exchange of letters concerning scientific work and, later, by a quasi-institutionalized arrangement embodied in the office of secretary or correspondent; finally, these were followed by formal journals containing the proceedings of meetings and other communications. The scientific enterprise of that era was small enough to permit reasonably adequate communication on the basis of the small number of journals and the occasional publication of books, supplemented by face-to-face interaction and correspondence by letter.

The so-called "communications explosion" in science is new primarily in the sense that it is now widely recognized as a problem. But as Price (1963) has pointed out, the amount of scientific publication has been growing exponentially, doubling every ten to fifteen years over the past three centuries. While there is no universal agreement, one of the more "conservative" estimates is that there are over thirty thousand scientific journals presently in existence and that there are more than a million papers published in them each year (Gottschalk & Desmond 1962; Bourne 1962). The sheer number of papers being produced annually has brought open recognition of a number of serious communication problems. Among the more important ones are the time lag between the completion of a paper

for publication and its appearance in a journal; the increasing difficulty of "keeping up with the literature", and the increasing difficulties in searching the literature and retrieving relevant information. For these as well as a number of other reasons centering on a growing awareness of the importance of science for national welfare, economic growth, and even survival, interest in the communications problems of science has mushroomed since the 1950s.

The communications explosion within science has given rise to concerted efforts to deal with the problem on many fronts, especially with the aid of new technological advances. Through the use of computers and a variety of other technical devices, efforts are being made to facilitate the storage and retrieval of information, and considerable progress is being made along these lines (see Stevens 1965). However, some questions have been raised about this approach in terms of the changing functions of scientific communication. While professional librarians, documentalists, editors, linguists, abstractors, mathematicians, and others now broadly characterized as information experts seek to improve the effectiveness of scientific communication largely by technical means, social scientists have begun to examine the social aspects of the scientific communications process.

### The study of scientific communication

Despite the growing interest in and support for studying the scientific communications system, there is still little systematic knowledge about it. A review of what is known must, therefore, be guided by a broad conceptual scheme rather than restricted to the specific questions now thought to be crucial for an understanding of the subject. The scheme has three major components: the *functions* of scientific communication, for scientists as well as for science generally; the various *channels* through which communications flow, the intervening variables or *situational factors* which influence the relationships between channels and functions.

**Functions of communication.** Menzel (1958) lists a number of functions performed by scientific communication: (1) providing answers to specific questions, (2) helping the scientist to stay abreast of new developments in his field, (3) helping him to acquire an understanding of a new field; (4) giving him a sense of the major trends in his field and of the relative importance of his own work; (5) verifying the reliability of information by additional testimony, (6) redirecting or broadening his span of interest and attention; and (7) obtaining critical response to his own work. He notes, too,

Merton's seminal discussion (1957) of the importance to scientists of professional recognition—a major reward for scientific achievement which is also carried by the communications system of science.

An understanding of these functions helps to pinpoint specific problems in the communications process and calls attention to the importance of the many different forms of communication behavior in which scientists engage. This is undoubtedly an important first step in going beyond the purely technical aspects of the storage and retrieval of scientific information, and studies concerned with these functions should lead not only to a better understanding of the communications process in science but also, it is hoped, may have practical utility in suggesting ways to improve the process.

One of the most important functions of communications for science as a whole is to provide a cumulative record of the "certified" knowledge which exists at any given point in time. Without such a record it is doubtful that science could continue to develop as a viable system. This record constitutes the point of reference for each scientist, providing him with the foundation from which he may make his own contributions toward extending what is known. This is not to imply that what is already in the record is immutable. Quite the contrary—the record is always subject to change in the light of new evidence, newly available techniques, and new discoveries. But whether a contribution is an extension of previously accepted knowledge or a new interpretation of what is already known, it is always and necessarily a matter of the record.

The current communications crisis raises questions about the nature of this record as well as about a number of widely held assumptions concerning the nature of the scientific communications process. For example, the fact that a paper has been published has usually implied that it has been reviewed by a jury of competent peers. However, there is at least the suspicion that the continued proliferation of journals has inevitably "watered down" the rigor of professional review. It is not even always possible to establish a simple correlation between the quality of any particular scientific paper and the reputation of the particular journal in which it appears. In any case, the quality of papers published is becoming an increasingly important problem. No matter how much retrieval procedures may be improved, the question of what is worth retrieving deserves much more attention.

Another major assumption has been that once a paper is published, not only is it accessible to the

scientific community, but also it will actually be read by the scientists concerned. However, the flood of publications has tended to undermine the accessibility of relevant papers, at least within the traditional communications practices. Not only may a single paper be buried in the flood of all papers published, but accessibility is also impeded by the growing trend toward increased specialization of journals, so that relevant papers may appear in sources not normally reviewed. Finally, the flood of publications has made it impossible for any one scientist to read more than a small fraction of what is potentially relevant. Some preliminary studies indicate that perhaps only about 25 scientists may actually read any particular paper which is published.

Further, an increase in the amount of information in the field tends to place greater strain on the integrative capacities of theory in that field, so that relationships among the contributions of different scientists become more difficult to determine. As a body of knowledge becomes "disorganized" in this way, the significance and even the validity of new contributions are more difficult to assess, and effective scientific communication may become a property only of small networks of scientists working on the same specific topics rather than of an entire discipline (Hagstrom 1965).

**Channels of communication.** The growing awareness of informal or relatively private channels of communication among scientists (U.S. President's . . . 1963; Schilling 1963–1964; Kaplan 1964) is but one aspect of the general problem of the different channels of communication and the different functions of each. The various channels of scientific communication are usually thought of as ranging from the most formal to the most informal, in terms of the degree to which the information flowing through them is codified and generally available to all scientists. It is perhaps more useful, however, to think of them as ranging from *structured* or planned channels which are known about in advance (such as journals, books, and even announced meetings) to *unstructured* channels through which information is accidentally acquired (such as finding useful information in the literature while searching for something else, or a casual conversation which yields unlooked-for information). A further complication must be introduced insofar as we are dealing with a highly fluid situation in which yesterday's unstructured channels become tomorrow's structured ones. The ease with which one can talk by telephone to a colleague three thousand miles away may make this a far more effective means of finding particular bits



of information than the ordinarily accepted techniques of searching the literature.

While precise knowledge of the proportion of information important to scientists acquired through various channels is still lacking, everything we know indicates that there is a much greater reliance on "accidental" and unstructured channels than has formerly been realized. Menzel (1958, p. 47) emphasizes that "it becomes imperative to consider the information network as a system. . . . What is little better than an accident from the point of view of an individual may well emerge as a predictable occurrence from a larger point of view."

While it is possible to try to match different channels with the different functions performed by scientific communication as discussed above, in practice such an exercise appears to be futile. Each of the functions listed can be, and apparently is, actually served by many different channels. Thus, it is entirely possible that the answer to a specific question can be found in an article in a published journal, in a preprint which arrived in the morning mail, during the course of a telephone conversation with a colleague about some other matter, or in a chance encounter in the corridors outside a scientific congress. A review article might be the best single source for providing a sense of the major trends in a field, but one might also get the same results, and often more quickly, from talking to a number of colleagues closely involved in the field. In the present circumstances, it seems safe to say that merely exploring one channel or source (especially a traditional one) is not necessarily the most effective way of treading through the maze of messages being communicated.

**Participation in structured channels.** The literature of science, primarily journals, constitutes the most important structured channel of communication within science; roughly two-thirds of the scientists studied by several investigators cited the journals as the most important single channel through which they learned of new developments in their fields of primary interest (Menzel 1958; 1960). To keep up with secondary fields, scientists typically turn first to recent textbooks and then to abstracts and review articles for guidance (Menzel 1958). The degree to which a body of knowledge is theoretically well-organized seems to influence the concentration of important information in specific channels; Menzel found chemists (a relatively well-organized field) reporting that two-thirds of the articles they read would be found in the three journals they listed as "most important," while zoologists (a relatively unorganized field) reported that only a quarter of their reading would

be found in three such journals. He found also that an average of 8.18 journals was needed to account for 75 per cent of the nominations of "three most important journals" by chemists, while the comparable figure for zoologists was 15.76 journals (1960).

Price (1963, table 2, p. 45) suggests that 6 per cent of the men in a field will produce half the published literature in that field and 2 per cent will produce about a quarter of it. Reasoning after Lotka (1926) that the number of scientists producing  $n$  papers is proportional to  $1/n^2$  (an inverse-square law of productivity), he calculates that the average scientist should produce about 3.5 papers during his working career. Meltzer (1956), however, found that U.S. physiologists had produced an average of four to five papers (including chapters in books and coauthored papers) within three years, and Schilling (1963-1964) seems to arrive at an even higher figure for the lifetime productivity of bioscientists.

Such figures are not directly related to the use of the literature; Price (1963) finds that the "half life" of a given article is about 15 years; that is, half of the articles cited by papers published in a given year will be less than 15 years old. Various other studies indicate that the half life in use is much shorter than this; more than half of the withdrawals from two technical libraries studied were less than five years old, and more than half of the "reading acts" by scientists in a U.S. government laboratory were devoted to materials less than two months old (Menzel 1960).

The amount of time a scientist devotes to the use of structured channels probably varies greatly with situational factors, although little is known about this at present. On the basis of some 25,000 random-time observations of 1,500 U.S. chemists, Halbert and Ackoff (1958) conclude that nearly 50 per cent of their subjects' time is spent in some form of communication and a third of it in specifically scientific communication. Roughly 10 per cent of the chemists' time is spent in general discussion, slightly less than this in receiving information orally, the same amount in reading unpublished materials, and about 5 per cent is spent in reading published materials. Estimates of the proportion of reading done for "specific uses" rather than for "general interest" vary from 20 to 80 per cent, and they are probably influenced heavily by the ways in which the data were gathered as well as by situational factors (Menzel 1960).

A number of studies report that the average amount of time per week spent in reading published materials is about five hours, but situational

factors are again highly important. Scientists in basic research seem to devote only half as much time to reading as do scientists in applied research; however, the former are much more concerned with archival literature, while the latter make greater use of unpublished literature (Menzel 1960). Törnudd (1958) suggests that physical isolation from channels of oral communication produces greater dependence among Danish and Finnish scientists upon published literature, and the rate at which a field is developing seems also to be important—the more rapid the advances, the greater the reliance upon relatively unstructured channels (Menzel 1958).

Schilling's study of bioscientists (1963–1964) found that age is unrelated to dependence on written, as compared with oral, channels; the bioscientists typically rate the latter as only half as important as the former. Moreover, women in the biological sciences seem slightly more dependent on the literature, probably because they have fewer opportunities to engage in unstructured communication: they visit other laboratories less often, hold fewer professional offices, and receive fewer preprints.

As the amount of published literature increases, scientists are turning to other structured channels as well as to unstructured channels. Orr (1964) notes a steady increase in the number of meetings and conferences held in the biomedical sciences each year since 1927, an increase in the attendance at such meetings, and a slight increase since 1947 in the percentage of research funds used for travel. Menzel's study of 76 scientists (1958) found that they attended an average of 2.5 meetings per year, and Garvey (see American Psychological Association . . . 1964) found that psychologists in an academic setting attended an average of three per year, while those in a government laboratory averaged about two. Yet Menzel notes that relatively few scientists admit obtaining significant information from the formal presentations at meetings and concludes that "the functions of scientific meetings are not those which ostensibly motivate the bulk of their programs, but other forms of communication—symposia, corridor meetings, the presence in one room of those interested in a single area . . ." (1960, vol. 1, p. 47).

**Unstructured communication.** Unstructured (sometimes called informal or unplanned) communication among scientists often provides specific information which a scientist knows he needs, but probably its major importance lies in providing him with useful information which he did not know existed. Scientists generally report between 65 and

90 per cent success in locating needed information in the literature (Menzel 1960; Törnudd 1958), but they obviously cannot estimate their success in obtaining information of which they are unaware.

Such unknown information varies from the specific (a new experimental technique) to the general (news that another individual is working on a particular problem), but it is almost always obtained, directly or indirectly, "by accident." Considering the usual delays in publication and the general difficulties involved in keeping abreast of the literature, it is presumptive evidence of the efficiency of unstructured channels that only about one scientist in five reports ever having received information "too late"—that is, information which would have influenced the course of his research had he received it sooner (Menzel 1960).

Menzel notes four major unstructured channels through which information reaches scientists: a scientist informs a colleague of his current interests and is given an item of pertinent information in return; a colleague conveys information which he knows the scientist will be interested in; a colleague volunteers the information while they are together for a different purpose; and he finds a useful item of information in the literature while searching for something else. In all but the last, the scientist is dependent on his colleagues, who will know of his needs and interests only if he tells them or if they are indicated by his previous publications. The effectiveness with which a scientist uses these channels, then, would seem to be related both to his ability to make his needs known and to the frequency with which he comes into contact with other scientists in his field.

**The flow of professional recognition.** Unstructured channels can operate effectively only so long as scientists feel the need to assist each other and share the values of "communism" and "disinterestedness" which encourage them to share their findings freely, without regard for personal gain (Merton [1949] 1957, pp. 556–560). Yet personal gain of an honorific nature is involved; Merton has referred to this as professional recognition (1957), and Storer (1966) has suggested that it is "competent response" to a contribution. If information flows in one direction through the various channels of scientific communication, response to it flows in the other. Anything which signifies the value, to one or more scientists, of information received from others—eponymy, prizes, election to professional office, footnotes, even personal thanks—serves to sustain the scientist's motivation through confirming the goodness of his work and his successful performance as a scientist. Glaser (1964) has docu-



mented some of the consequences for the motivation and career plans of scientists when there is a lack of "adequate recognition," but work on the various channels through which recognition flows has barely begun (Kaplan 1965a).

### Improving scientific communication

The extent and success of efforts to speed and to make more effective the dissemination of information in science vary greatly by fields. In physics, the weekly *Physical Review Letters* has made it possible to bring brief announcements of recent findings to readers, usually within a month of submission, and there is a variety of other newsletters, data-card services, and regular announcements of work in progress which are now being established in different fields. The American Psychological Association's Project on Scientific Information Exchange in Psychology represents a major effort to develop and then apply new information to the improvement of communication within the field of psychology.

Another technique, facilitated by the use of high-speed computers, is the citation index, which enables one to trace the influence of a given paper forward in time; this will apparently be of value to historians and sociologists of science as well as to those concerned with the substantive content of the materials cited (Institute for Scientific Information 1964). The suggestion that archival publication be partially replaced by central depositories, from which materials may be acquired on request after learning of them through title lists and abstracts, has not yet met with success; the obstacles to its adoption apparently lie more in the desire of scientists to be assured that their contributions will go to a "guaranteed" audience (as when published in a journal) than in the technical problems involved (U.S. President's . . . 1963).

Suggestions have also been made toward providing greater opportunity for scientists to make use of unstructured channels of communication: encouraging attendance at meetings and visits to other institutions, arranging teaching duties so as to leave some days free for travel, and allowing more time at announced meetings for discussion sessions (Menzel 1958). Such suggestions are basically concerned with increasing the amount of personal contact among scientists, and their success will probably be contingent upon the amount of funds available for such purposes.

It may be predicted that the study of scientific communication will become increasingly important as the difficulty of disseminating information widely and rapidly mounts. The field requires much

more work in conceptualizing the nature of the communications network and its relation to the social structure of science, as well as in the collection of more data. While much of the concern expressed today about problems of scientific communication focuses on developing new techniques for the storage and retrieval of information, and while some concern is focused on the social aspects of the communications process, very little attention has been devoted to the underlying problem of what should be communicated and in what form.

Finally, there is little doubt that each scientist will have to bear a greater responsibility for the transfer of information and not leave it largely to the professional documentalist. As a recent analysis of the problems of scientific communication (U.S. President's . . . 1963, p. 1) notes, "The technical community generally must devote a larger share than heretofore of its time and resources to the discriminating management of the ever increasing technical record. Doing less will lead to fragmented and ineffective science and technology."

NORMAN KAPLAN and NORMAN W. STORER

[See also CONFERENCES; DIFFUSION, article on INTERPERSONAL INFLUENCE; INFORMATION STORAGE AND RETRIEVAL.]

### BIBLIOGRAPHY

- AMERICAN PSYCHOLOGICAL ASSOCIATION, PROJECT ON SCIENTIFIC INFORMATION EXCHANGE IN PSYCHOLOGY 1964 *The Discovery and Dissemination of Scientific Information Among Psychologists in Two Research Environments*. American Psychological Association, Project on Scientific Information Exchange in Psychology, Report No. 11.
- American Psychologist 21, no. 11. → This issue (November, 1966) contains eight articles on scientific communication, particularly in the field of psychology.
- BARBER, BERNARD; and HIRSCH, WALTER (editors) 1962 *The Sociology of Science*. New York: Free Press.
- BOURNE, CHARLES P. 1962 *The World's Technical Journal Literature: An Estimate of Volume, Origin, Language, Field, Indexing, and Abstracting*. *American Documentation* 13:159-168.
- GLASER, BARNEY G. 1964 *Organizational Scientists: Their Professional Careers*. Indianapolis, Ind.: Bobbs-Merrill.
- GOTTSCALK, C. M.; and DESMOND, W. F. 1962 *Worldwide Census of Scientific and Technical Serials*. Paper delivered at the annual meeting of the American Documentation Institute. Unpublished manuscript.
- HAGSTROM, WARREN O. 1965 *The Scientific Community*. New York: Basic Books.
- HALBERT, MICHAEL H.; and ACKOFF, RUSSELL L. 1958 *An Operations Research Study of the Dissemination of Scientific Information*. Pages 87-120 in *International Conference on Scientific Information, 1958, Preprints of Papers for the International Conference on Scientific Information, Washington, D.C., November 16-21, 1958*. Area 1. Washington: National Academy of Sciences-National Research Council.

- INSTITUTE FOR SCIENTIFIC INFORMATION 1964 *The Use of Citation Data in Writing the History of Science*, by Eugene Garfield et al. Philadelphia: The Institute. → Constitutes a report of research for the U.S. Air Force Office of Scientific Research under contract AF 49(638)-1256.
- INTERNATIONAL CONFERENCE ON SCIENTIFIC INFORMATION, 1958 1958 *Preprints of Papers for the International Conference on Scientific Information*, Washington, D.C., November 16-21, 1958. Areas I-viii. Washington: National Academy of Sciences-National Research Council.
- KAPLAN, NORMAN 1964 *Sociology of Science*. Pages 852-881 in Robert E. L. Faris (editor), *Handbook of Modern Sociology*. Chicago: Rand McNally.
- KAPLAN, NORMAN 1965a *The Norms of Citation Behavior: Prolegomena to the Footnote*. *American Documentation* 16:179-184.
- KAPLAN, NORMAN (editor) 1965b *Science and Society*. Chicago: Rand McNally.
- LITTLE [ARTHUR D.] INC. 1959 *Basic Research in the Navy*. 2 vols. Cambridge, Mass.: Arthur D. Little, Inc.
- LOTKA, ALFRED J. 1926 *The Frequency Distribution of Scientific Productivity*. *Journal of the Washington Academy of Sciences* 16:317-323.
- MELTZER, LEO 1956 *Scientific Productivity in Organizational Settings*. *Journal of Social Issues* 12, no. 2: 32-40.
- MENZEL, HERBERT 1958 *The Flow of Information Among Scientists: Problems, Opportunities and Research Questions*. Unpublished manuscript, Columbia Univ., Bureau of Applied Social Research.
- MENZEL, HERBERT 1960 *Review of Studies in the Flow of Information Among Scientists*. 2 vols. in 1. Unpublished manuscript, Columbia Univ., Bureau of Applied Social Research.
- MERTON, ROBERT K. (1949) 1957 *Social Theory and Social Structure*. Rev. & enl. ed. Glencoe, Ill.: Free Press.
- MERTON, ROBERT K. 1957 *Priorities in Scientific Discovery: A Chapter in the Sociology of Science*. *American Sociological Review* 22:635-659.
- ORR, RICHARD H. 1964 *Communication Problems in Biochemical Research: Report of a Study*. *Federation Proceedings* 23:1117-1132.
- PRICE, DEREK J. DE SOLLA 1963 *Little Science, Big Science*. New York: Columbia Univ. Press.
- SCHILLING, CHARLES W. 1963-1964 *Informal Communication Among Bioscientists*. Parts 1-2. Unpublished manuscript, George Washington Univ., Biological Sciences Communication Project.
- STEVENS, MARY E. 1965 *Automatic Indexing: A State-of-the-art Report*. National Bureau of Standards Monograph No. 91. Washington: Government Printing Office. → Includes a bibliography of 662 items on pages 183-220.
- STORER, NORMAN W. 1966 *The Social System of Science*. New York: Holt.
- TÖRNQVIST, ELIN 1958 *Study on the Use of Scientific Literature and Reference Services by Scandinavian Scientists and Engineers Engaged in Research and Development*. Pages 9-65 in *International Conference on Scientific Information, 1958, Preprints of Papers for the International Conference on Scientific Information*, Washington, D.C., November 16-21, 1958. Area I. Washington: National Academy of Sciences-National Research Council.

U.S. PRESIDENT'S SCIENCE ADVISORY COMMITTEE 1963 *Science, Government, and Information: The Responsibilities of the Technical Community and the Government in the Transfer of Information; a Report*. Washington: Government Printing Office.

## SCIENTIFIC EXPLANATION

There is little disagreement that explanation is a major aim of science; there is much about the conditions that a proposed explanation must satisfy. Most of this disagreement, especially in history and in the social sciences, concerns the so-called nomological and deductive models of satisfactory explanation. There is also a more general controversy: many—perhaps most—scientists and philosophers maintain that these models clearly apply to the natural sciences and should apply to the social sciences as well, while others insist that they do not apply to any science, natural or social. For the most part, however, I shall bypass this controversy and discuss the applicability of the two models to the social sciences.

But first let us raise the logically prior question "What is an explanation?" There are three traditional answers to this: (1) to explain is to remove perplexity; (2) to explain is to change the unknown to the known; (3) to explain an event or type of event is to give its causes.

The first two answers have the virtue of reminding us of the psychological conditions that frequently obtain when we request an explanation. But this virtue is, in a way, the defect of these two approaches: they are too psychological, in the sense that they do not stipulate the logical conditions that must obtain before we have a satisfactory explanation. Clearly, not every way of removing perplexity can count as an explanation. Again, it should be noted that both of these approaches fail to point out that when we explain an event or type of event, we attempt to relate it to others like it and to give a systematic account of phenomena which will show how they are interrelated. In short, these answers fail to emphasize the systematic interconnection between explanation and theory construction.

The third answer does not share in these defects. The thesis that to explain an event is to give its cause is one that goes back to Aristotle and has been defended at length by Mill (1843) and other distinguished logicians. But this approach, whatever its virtues, cannot count as an adequate theory of explanation. At best it is a partial account of the explanation of events, but not of regularities, dispositions, and other types of phenomena



which we try to explain. Moreover, since the term "cause" is not in itself a very clear one, a theory which builds on it without attempting any explanation of it is not a reliable guide. However, many philosophers and social scientists have been convinced by the arguments offered by Hume and his followers that a causal statement entails a lawlike one, i.e., that a statement of the form "Event A caused event B" entails a statement of the form "Events like A regularly precede and are contiguous with events like B." If this analysis be granted, then the approach to explanation which emphasizes causation is compatible with the deductive and nomological models.

**Nomological and deductive models.** Supporters of the nomological model of scientific explanation insist that no event or regularity is explained by a proposed explanation unless it contains a scientific law (*nomos* is Greek for law). Supporters of the deductive model accept this but go further; they maintain that an event or regularity is explained if a sentence describing that event or regularity is deduced from a set of premises containing (a) a scientific law and (b) any other premises that may be required to make the deduction logically correct (provided, of course, that these premises are true).

Thus, given a sentence *P* that can be used to describe an event, the nomological model states that the event is not explained by a series of sentences  $S_1, S_2, \dots, S_n$  unless this series contains a scientific law. To this the deductive model adds that if  $S_1, S_2, \dots, S_n$  contains one or more laws, and if *P* can be deduced from  $S_1, S_2, \dots, S_n$ , then  $S_1, S_2, \dots, S_n$  is a satisfactory conjunction of explaining sentences. This assumes that *P* stands for a sentence describing a particular event. When we want to explain not an event, but a trend or a regularity, all the sentences needed, both descriptive and explanatory, would of course be far more complex, but the over-all form of the explanation would stay the same.

On occasion some defenders of the deductive or the nomological models add that the explaining premises  $S_1, S_2, \dots, S_n$  must contain or constitute a theory. But this addition does not clarify matters at all, because the meaning of the term "theory" is itself unclear. Sometimes the term is applied to any systematically interconnected set of sentences; sometimes it appears to mean the same as "hypothesis"; sometimes it is used interchangeably with "law." Clearly, if "theory" and "law" are to be used interchangeably, acceptance of either the deductive or the nomological model requires us to believe that when we explain, we appeal to or use a theory. Since the term "theory" has so many

uses, it is perhaps best to avoid using it in this context and to say simply that the premises of nomological and deductive explanations contain laws.

Candor demands, however, the admission that we are far from a completely satisfactory statement of either model. Thus, in the nomological model the meaning of "law" is just as much in dispute as that of "theory," while in the deductive model the kinds of sentences that may appear in the series  $S_1, S_2, \dots, S_n$  need to be specified in greater detail. On the other hand, there is little point in being concerned with such details if this whole approach to explanation is not valid. To see whether it is, I shall try to remove some misunderstandings about the nature of the nomological and deductive models, and in so doing to counter some of the arguments against them and admit the truth of others, at least in qualified form.

Do the models involve determinism? Both the deductive and the nomological models have been criticized on the grounds that they involve commitment to determinism. But this criticism is irrelevant. One may accept either model and still be agnostic about the thesis that every event is covered by laws—a thesis frequently given as one specification of determinism. Thus, philosophers who accept this thesis frequently argue that any sentence truly descriptive of an event is in principle deducible from some series of sentences containing one or more laws. But clearly this is not the same as claiming with the supporters of the nomological and deductive models, that if a sentence truly descriptive of an event is deduced from a series of sentences containing a law, then the event is explained.

Of course, many defenders of the nomological and deductive models do accept an even stronger version of determinism than the thesis that every event is covered by laws. But they often accept it as a "guiding principle of inquiry," to use C. S. Peirce's phrase. Thus, they may insist that the scientist should not merely seek a variety of general laws for explaining various events and regularities but should construct and confirm one general theory that would explain all events and regularities, or at least all those within the purview of a particular science. But obviously a guiding principle is not the same thing as an article of faith, and many supporters of the nomological and deductive models agree that there may very well be no such general theory or theories for scientists to discover.

Can social scientists use the models? The deductive and nomological models are sometimes

criticized on the grounds that social scientists have not succeeded in confirming any laws and therefore cannot use either model. To this it can be objected that some social scientists at least claim to have discovered laws, and that in any case only true or completely acceptable explanations, of which social scientists admit to having provided very few, are supposed to conform to the models' requirements.

But though the criticism as it stands is irrelevant, it does indicate that it is not enough to consider only true explanations or fully acceptable ones; we must consider not only the truth of proposed explanations but also their degrees of acceptability. Clearly, when a scientist proposes an explanation and employs a generalization, he obviously does not know that the generalization is true and, hence, that he is in possession of a law. The nomological and deductive models must therefore be supplemented, as follows. Let us use the term "lawful statement" to denote any highly confirmed lawlike statement—that is, any statement that (a) would be a law if it were true and (b) has in fact been found to hold true on numerous occasions. Explanations containing such statements have, other things being equal, a high degree of acceptability, and many of them also turn out to be true explanations—that is, the lawful sentences that they contain are true. But, of course, explanations that at one time were highly acceptable because the lawlike sentences they contained were at that time highly confirmed may turn out to be false (a classic case of this is Newtonian mechanics). This distinction between explanations that are acceptable at a given time and true explanations that simply conform to the deductive model is not *ad hoc*. Rather, it corresponds to the distinction, usually applied to statements, between "true" and "well confirmed."

Here it may be objected that social scientists do not refer to laws and, hence, that it is misleading to appeal to laws when discussing the logic of social scientific explanation. But this objection is a weak one. Social scientists tacitly employ many laws that are so well known that no special reference to them is needed. Furthermore, it should be emphasized that many social scientists do in point of fact refer to laws—economists, for instance, to Say's law of markets, and sociologists to Michels' "iron law of oligarchy." It may also be objected that the distinction between a highly confirmed generalization and a lawlike one is artificial. But once again the answer is readily available. Thus, in everyday life we would not treat the generalization "All American presidents are Christian" as lawlike

or as a candidate for a law, even though all the evidence to date for it is positive and, in that sense, it is well confirmed. Observe further that the distinction we are discussing here is parallel to, although admittedly identifiable with, the frequently made distinction between causal statements, on the one hand, and statements of mere association, or statements of correlations which may be spurious, on the other.

What, then, are we to make of proposed explanations that do not contain lawful statements? We certainly do not always know on what law, if any, an investigator has based his explanation of certain data, even if we are inclined to accept what he says. As an argument against the nomological and deductive models this is not a very strong one, since most social scientists would admit that such proposed explanations should be replaced by more carefully stated ones. Besides, if the acceptability of such an explanation depends, as it usually does, on the degree to which it approximates a nomological or a deductive explanation, then it cannot be called an alternative to these latter, but merely an inferior substitute for them.

**Motivational explanations.** It is a commonplace that social scientists offer motivational explanations, but it is not at all obvious how such explanations should be interpreted from a logical point of view. Note first that when a motivational explanation is offered, there is no necessary commitment to the thesis that there are private internal entities called motives which are responsible for behavior or action; all that is involved is some reference to the wants, preferences, or aims of the individual or group whose action is being described or explained. In this sense all Freudian theory, for instance, makes tacit or explicit reference to motives, and so do all explanations by economists that refer to the utility of the agent or group whose behavior is being discussed. Similarly, tacit reference is made to the aims of a person when sociologists or social psychologists appeal, in order to make sense of his behavior, to the reference group with which he identifies himself. What then are we attributing to a person when we attribute a motive to him? The answer can readily be suggested: we are attributing to him a disposition to behave in a certain way or a disposition to prize certain outcomes over others. But this constitutes another possible objection to the nomological and deductive models. When a motivational explanation is offered, no appeal is made to a general law about the way human beings behave but, rather, to a dispositional statement about an individual. But it is very difficult to give the exact grounds upon which we can distinguish



between laws and dispositional statements. Admittedly, it seems possible to do without general laws if we restrict our attention to those cases in which the action seems to have been inspired by only one motive and in which there is no need, for explanatory purposes, to consider what the agent thought about the alternative ways available to him for the satisfaction of his motive. But when we consider the more usual cases of action—action performed out of many motives and action after deliberation about alternatives—it will not do to consider only the agent's disposition. Can we handle both these latter cases by reference to law? The answer is not clear. Some social scientists have thought it plausible to assume, as at least a first approximation, that insofar as people are rational, they try to maximize expected utility. But many suspect that the generalization tacitly appealed to here is a disguised tautology. Nevertheless, the defense of this generalization, or of alternatives to it, indicates that in order to explain actions for which there are many motives, some social scientists do look for laws and, hence, act in accord with the deductive model.

Others suggest that we should not seek general laws but construct ideal types instead [see *TYPOLOGIES*]. To construct an ideal type explanation for an action *A* by a person *P* is (1) to impute at least one motive, *T*, to *P*; (2) to list the alternatives confronted by *P*; (3) to show that an action of the same type as *A* is a rational one—perhaps the only rational one—for *P* to perform, given *T*; (4) to show that performing *A*, given alternatives  $A_1, A_2, \dots, A_n$ , is an efficient way of satisfying *T*. If, given the same motive and set of alternatives, *P* does not perform *A*, defenders of the ideal type approach would suggest that we must institute a search for the factors making for irrationality on the part of *P*.

One argument against this version of the ideal type model of explanation is that the motive imputed to the agent may in fact be the wrong one; another is that the model is incomplete unless the terms "rational" and "efficient" are explicated. Moreover, ideal type explanation sometimes appears to depend on a general law about the behavior of rational agents and, hence, to approximate either the deductive model or, at least, the nomological one.

**Functional explanations.** Conclusions similar to those drawn above can, I think, be applied to so-called functional explanations, for these either lack *prima-facie* acceptability or, when they have it, are not evidently at variance with the deductive or nomological models. Thus, a common type of functional "explanation" may state merely that a certain institution plays an indispensable role in

a society. Such a statement provides us with nothing more than a necessary condition and, hence, lacks *prima-facie* acceptability as an explanation. However, the stronger types of functional explanation do show that, holding certain things constant, the presence of a certain mechanism is a sufficient condition for the existence of a certain state of affairs. In my opinion, explanations of this type conform to the deductive model, for to say that *S* is a sufficient explanatory condition for *T* is, I think, equivalent to saying that the statement "If *S*, then *T*" is not only true but also lawful.

Of course, we need more complex laws than the ones thus far discussed if we are to provide descriptions and explanations of events in systems that are purposive and self-corrective. But it is far from obvious that this requires a special kind of explanation called teleological, or functional, explanation, and even less obvious that societies—at least complex ones—can be analyzed in teleological terms. Certainly, societies are not totally purposive, nor do such purposes as they may sometimes be said to have (making war, for example) always remain the same. At most, societies may tend to reorganize themselves in order to keep certain properties constant. But it is hard to say which properties these are [see *SYSTEMS ANALYSIS, especially the article on SOCIAL SYSTEMS*].

**Causal explanation.** It has been maintained at least since Aristotle that to explain an event is to give its causes. But this does not necessarily involve controverting the nomological or the deductive models, since it can be argued that to give the causes for an event is to present a deductive explanation containing a causal law. Given this approach, we may then try to explicate the uses of the term "causal law."

A statement is occasionally called a causal law if it specifies either a sufficient condition or necessary and sufficient conditions for a certain type of occurrence. In logic, such a statement may take any of these forms: (1) if *A* then *B*; (2) *A* if and only if *B*; (3) *A* is a function of *B* (when *A* and *B* stand in the place of terms designating measurable properties). Deterministic laws, which describe how a system, with states that are described mathematically, evolves over time, are also called causal laws. Finally, the term "causal" is applied to laws of the form "if *A* then *B*" when (1) *A* denotes a type of event that comes just before an event of type *B*; (2) *A* and *B* are events or episodes in bodies or agents that are spatially contiguous; (3) the occurrence of *A* is to be considered a sufficient condition for the occurrence of *B* even though an event of type *B* might occur without one of type *A*.

preceding it. Stimulus-response generalizations frequently meet these three criteria—criteria that represent perhaps the most legitimate use of the term “causal law.”

Not all explanations that appeal, whether overtly or tacitly, to the notion of cause make use of causal laws in any of the senses discussed above. Accordingly, it is not evident that all causal explanations should conform to the nomological model, and still less to the deductive. Nevertheless, the view I am advocating has good philosophical precedent. At least since Hume many philosophers have argued that to assert of an event, *A*, that it caused another event, *B*, is equivalent to appealing tacitly to a law of the form “Events of type *A* are sufficient conditions of events of type *B* and precede them.”

This view is not completely persuasive. Often we say that an event *A* caused an event *B* when we mean only that the event *A* was a necessary condition for the event *B*. Also, we sometimes refer to an event *A* as the cause or a cause of an event *B* when we mean to assert that *A* was either a sufficient or a necessary condition for the occurrence of *B* only when conjoined with other such conditions. Moreover, when we say that *A* caused *B*, we are often unprepared to specify a general law relating events of type *A* to events of type *B*. Thus, we might be convinced that a given remark had caused the hearer to blush, but we might not be ready at the same time to cite a general law about the relation between remarks and blushing. Nevertheless, it does seem that when we specify a cause for an event and believe that this specification provides us with an explanation of the event, we tacitly commit ourselves to some generalization, however vague. Thus, to revert to the illustration given above, a person who insists that *X* blushed because *Y* said something might be challenged by *Z* to the effect that *X* didn't blush the last time he heard a similar remark, and that therefore the present remark couldn't have been responsible for his blushes. But this observation, though it helps to diminish the distance between explanations that contain causal laws and explanations that merely appeal to the notion of cause, does not abolish that distance completely. Whether it can ever be abolished is still under debate.

**Statistical explanation.** The status of statistical explanation, and its relation to nomological and deductive explanation, are topics subject to much philosophical controversy.

Statistical explanations are explanations containing lawlike statements based on observation of statistical regularities and/or on the statistical theory of probability. It is generally accepted that probabilistic statements, as interpreted in the sci-

ences, cannot be finally confirmed or even disconfirmed by observational evidence. It is also accepted that we cannot deduce from any statistical generalization a statement to the effect that any particular event must occur. [See PROBABILITY.]

It is not so widely recognized, however, that the problem of statistical explanation is much harder to deal with in the case of statistical generalizations that are either spatially or temporally restricted. Thus, if we knew that 90 per cent of the people in Milwaukee are Democrats and that 85 per cent of all living graduates of Yale University are Republicans, we could not use these generalizations as they stand to explain the voting habits of Yale graduates living in Milwaukee. It is therefore most important, when discussing statistical explanations, to appeal to the nomological, and not the deductive, model, for clearly we cannot expect to construct deductive explanations on every occasion when we use statistical generalizations for explanatory purposes. However, it would be a mistake to think that we can never have a deductive explanation under these circumstances, for we may deduce a statistical generalization from a conjunction of two or more other statistical generalizations.

It is important to note that we have not discussed the very complex but relatively typical situation that arises when social scientists who are using statistical data in an attempt to distinguish between the effects of several factors note that one factor is more relevant than the others for the purpose of explaining the phenomenon. Discussion of this and other such situations has been omitted because their logic is under much dispute. Many of the issues involved have been considered by Nagel (1952), though for some of the problems that are raised by statistical data and their analysis, the interested reader should consult works by Blacklock (1964) and Boudon (1965; see also Ando et al. 1963).

**Amplifying the models.** We have discussed some objections to the nomological and deductive models; it is now time to consider amplifying them in order to meet these objections. Let us begin by emphasizing the trivial point that we never explain an event as such, but only selected aspects of it. In other words, it is not the event itself that is explained but the event under a given description. Thus, if a person sits down on a tack and yells, we may say that he yelled because he sat on the tack. But we obviously are not ready to claim that we can explain the specific pitch or duration of the yell in the same terms.

Observe, however, that an explanation may be totally satisfactory once we specify what aspect of the event we are trying to explain. This is of special



relevance to the discussion of historical explanation. Historians have often noted that any event can be described in many ways and that some explanations of events are satisfactory only if the events are described schematically. Thus, we might be able to explain that a group migrated because it hoped to improve its lot but not be able to explain why it migrated on a specific date and to a specific country. Some historians, noting that there seems to be no end to the detail in which a given event can be described, have concluded that, for this reason, no event can ever be explained, nor can the causes of an event be fully specified. Against this view, I would like to urge that we discuss not the explanation of an event but the explanation of an event under a given description. This would allow us to accept some explanations as totally satisfactory even if we wanted to replace them with explanations of the same event under a more refined description.

Finally, it should be noted that in order to give an explanation of an event of a certain type, it is not necessary to cite laws or theories about events of that type. This is often overlooked in discussion of how to explain certain social phenomena; for instance, many social scientists seem to believe that in order to explain instances of crime or divorce, they need a general theory of crime or divorce. But this is dubious. To explain why a man slipped on a banana peel, we do not need a general theory of slipping. Rather, the laws that we need to cite are general laws in which terms like "slipping" do not occur. The essential point is that in order to explain an event, we often must redescribe it in terms of a given theory; moreover, two events that in ordinary life are both classified under the same general term may, for purposes of explanation, have to be described differently and explained by different general theories. To take an everyday example: A man who hurries to meet a friend and a man who flees to avoid disaster may both be running, yet the explanations for their behavior are different. More complicated examples, requiring explanations based on abstract theories, would require that the events needing explanation be redescribed in terms drawn from those theories. Although this type of redescription was not required in our example, the point remains that two events ordinarily classified as instances of the same type may have different explanations. The same applies to explanations involving not events but groups, epochs, social systems, and societies; we must specify what aspects of these entities we want to have explained.

SIDNEY MORGENBESSER

[Directly related are the entries CAUSATION; EXPERIMENTAL DESIGN; FUNCTIONAL ANALYSIS, article on VARIETIES OF FUNCTIONAL ANALYSIS; PREDICTION; SCIENCE, article on THE PHILOSOPHY OF SCIENCE. Other relevant material may be found in HISTORY, article on THE PHILOSOPHY OF HISTORY; MOTIVATION; MULTIVARIATE ANALYSIS; POSITIVISM; STATISTICS, DESCRIPTIVE; and the biographies of COHEN; PEIRCE.]

#### BIBLIOGRAPHY

- ANDO, ALBERT; FISHER, F. M.; and SIMON, HERBERT A. 1963 *Essays on the Structure of Social Science Models*. Cambridge, Mass.: M.I.T. Press. → See especially pages 5-31 and 107-112.
- ARGYLE, MICHAEL 1957 *The Scientific Study of Social Behaviour*. London: Methuen.
- BLALOCK, HUBERT M. JR. 1964 *Causal Inferences in Nonexperimental Research*. Chapel Hill: Univ. of North Carolina Press.
- BOUDON, RAYMOND 1965 A Method of Linear Causal Analysis: Dependence Analysis. *American Sociological Review* 30:365-374.
- BRAITHWAITE, R. B. 1953 *Scientific Explanation*. Cambridge Univ. Press.
- BROWN, ROBERT R. 1963 *Explanation in Social Science*. Chicago: Aldine.
- BUNGE, MARIO 1959 *Causality: The Place of the Causal Principle in Modern Science*. Cambridge, Mass.: Harvard Univ. Press. → A paperback edition was published in 1963 by World.
- CHRIST, CARL F. 1966 *Econometric Models and Methods*. New York: Wiley. → See especially the Foreword by Jacob Marshak.
- COHEN, MORRIS R.; and NAGEL, ERNEST 1934 *An Introduction to Logic and Scientific Method*. New York: Harcourt. → The first part of the original text was published in a paperback edition in 1962 by Harcourt, under the title *An Introduction to Logic*.
- DRAY, WILLIAM 1957 *Laws and Explanation in History*. Oxford Univ. Press.
- FEIGL, HERBERT; and MAXWELL, GROVER E. (editors) 1962 *Scientific Explanation, Space, and Time*. Minnesota Studies in the Philosophy of Science, Vol. 3. Minneapolis: Univ. of Minnesota Press. → See especially pages 231-273, "Explanation, Prediction and 'Imperfect Knowledge,'" by May Brodbeck.
- FODOR, JERRY A.; and KATZ, JERROLD J. (editors) 1964 *The Structure of Language: Readings in the Philosophy of Language*. Englewood Cliffs, N.J.: Prentice-Hall.
- GIBSON, QUENTIN B. (1960) 1963 *The Logic of Social Inquiry*. New York: Humanities.
- GREENWOOD, ERNEST 1945 *Experimental Sociology: A Study in Method*. New York: Columbia Univ. Press.
- HEMPEL, CARL G. (1952) 1963 Typological Methods in the Social Sciences. Pages 210-230 in Maurice A. Natanson (editor), *Philosophy of the Social Sciences: A Reader*. New York: Random House.
- HEMPEL, CARL G. 1965 *Aspects of Scientific Explanation, and Other Essays in the Philosophy of Science*. New York: Free Press.
- KAPLAN, ABRAHAM 1964 *The Conduct of Inquiry: Methodology for Behavioral Science*. San Francisco: Chandler.
- LOWE, ADOLPH 1965 *On Economic Knowledge*. New York: Harper.

- MADDEN, EDWARD H. 1962 *Philosophical Problems of Psychology*. New York: Odyssey. → See especially Chapter 3.
- MILL, JOHN STUART (1843) 1961 *A System of Logic, Ratiocinative and Inductive: Being a Connected View of the Principles of Evidence and the Methods of Scientific Investigation*. London: Longmans. → See especially Book 6.
- MORGENBESSER, SIDNEY 1966 Is It a Science? *Social Research* 33:255-271.
- NAGEL, ERNEST (1952) 1953 The Logic of Historical Analysis. Pages 688-700 in Herbert Feigl and May Brodbeck (editors), *Readings in the Philosophy of Science*. New York: Appleton. → First published in *Scientific Monthly*.
- NAGEL, ERNEST 1961 *The Structure of Science: Problems in the Logic of Scientific Explanation*. New York: Harcourt. → See especially pages 582-588.
- POINCARÉ, HENRI (1908) 1952 *Science and Method*. New York: Dover. → First published in French.
- POPPER, KARL R. (1934) 1959 *The Logic of Scientific Discovery*. New York: Basic Books. → First published as *Logik der Forschung*.
- REICHENBACH, HANS 1951 Probability Methods in Social Science. Pages 121-128 in Daniel Lerner and Harold D. Lasswell (editors), *The Policy Sciences: Recent Developments in Scope and Method*. Stanford Univ. Press.
- SCHROEFFLER, SIDNEY 1955 *The Failures of Economics: A Diagnostic Study*. Cambridge, Mass.: Harvard Univ. Press.
- SEGERSTEDT, TORGNY T. 1966 *The Nature of Social Reality*. Copenhagen: Munksgaard.
- SKINNER, B. F. 1953 *Science and Human Behavior*. New York: Macmillan.
- TOULMIN, STEPHEN E. 1953 *The Philosophy of Science: An Introduction*. London: Longmans. → A paperback edition was published in 1960 by Harper.
- WINCH, PETER 1958 *The Idea of a Social Science and Its Relation to Philosophy*. London: Routledge; New York: Humanities.
- WOLMAN, BENJAMIN B.; and NAGEL, ERNEST (editors) 1965 *Scientific Psychology*. New York: Basic Books. → See especially "Concerning an Incurable Vagueness in Psychological Theories," by Benbow F. Ritchie, and "Mathematical Learning Theory," by Richard C. Atkinson and Robert C. Calfe.

## SCIENTIFIC MANAGEMENT

See ADMINISTRATION, article on THE ADMINISTRATIVE FUNCTION; AUTOMATION; BUSINESS MANAGEMENT; OPERATIONS RESEARCH.

## SCREENING AND SELECTION

Screening and selection procedures are statistical methods for assigning individuals to two or more categories on the basis of certain tests or measurements that can be made upon them. Of concern usually is some desired trait or characteristic of the individuals that cannot be measured directly. All that can be done is to obtain an estimate for each individual from the results of the available tests

and then to make the assignment on the basis of these estimates. The central statistical problem is to evaluate the properties of alternative schemes for utilizing the available data to make the assignments in order to choose the scheme that best achieves whatever objectives are considered to be most relevant for the particular application.

For example, some educational selection schemes may be regarded in this light. The individuals might be high school students and the categories "admit to college" and "do not admit to college." The desired trait is future success in college, but only tests at the high school level are available. (As will be seen, many examples of selection and screening are somewhat more complex than this, particularly in their use of more than one level of screening. In this educational context one might instead use three categories: "admit," "put on waiting list," and "do not admit.")

Denote by  $N$  the number of individuals to be assigned and let  $c$  be the number of categories. For any individual, let  $Y$  denote the unknown value of the desired trait and let  $X_1, \dots, X_p$  denote the measurements or scores that can be obtained and used as predictors of  $Y$ . The screening or selection procedure is a scheme that specifies in terms of  $X_1, \dots, X_p$  how each individual is to be assigned to one of the  $c$  categories.

The screening may be done at one stage—that is, all the measurements  $X_1, \dots, X_p$  become available before an individual is assigned—or it may be multistage. The advantage of a multistage procedure is that it may allow some individuals to be assigned at an early stage or, at least, to be eliminated from contention for the categories of interest, thus permitting the resources available for performing the tests to be concentrated on fewer individuals in the later stages.

The terms *screening* and *selection* are largely synonymous, although in particular applications one or the other may be preferred. Sometimes, in order to avoid the possible connotation that certain categories may be more desirable than others, a neutral term such as *allocation* is used. The term *classification* has a different shade of meaning, referring to the identification of which of several distinct distributions each individual belongs to (In taxonomy, for example, classification involves the assignment of an organism to its proper species) [see MULTIVARIATE ANALYSIS, article on CLASSIFICATION AND DISCRIMINATION].

Various formulations of screening and selection problems have been proposed and investigated. A bibliography containing more than five hundred references has been given by Federer (1963). Of



special interest is the case  $c = 2$ , in which the individuals are separated into two categories, a selected group and the remainder; the success of the screening procedure is judged by the values of  $Y$  for the individuals in the selected group. This article will be primarily devoted to this case of two categories.

### The case where $N$ is large

This section deals with the case of separation into two categories where either the number,  $N$ , of individuals is large enough for their  $Y$ -values to be considered as forming a continuous distribution or, alternatively, the  $N$  individuals themselves are considered as a random selection from a conceptually infinite population. The object of the screening procedure is to produce a distribution of  $Y$ -values in the selected group that is, in an appropriate sense, an improvement on the original distribution. For example, one might want the selected group to have as high a median  $Y$  as possible.

In many applications, the feature of the distribution of  $Y$  that is considered most important is its mean. For example, in a plant-breeding program,  $Y$  might stand for the crop yield of the individual varieties, and the purpose of the program might be to select a set of varieties whose mean  $Y$  is as high as possible. The difference in the mean  $Y$  for the selected varieties from that of the original group is referred to as the "advance" or the "gain due to selection."

Cochran (1951) summarized the mathematical basis for selection procedures designed to maximize the mean  $Y$  in the selected set. He showed that the optimum selection rule to use at each stage should be based on the regression of  $Y$  on the  $X$ 's that are known at that stage. In the case where the joint distribution of these variates is multivariate normal, this regression is the linear combination of the  $X$ 's that has maximum correlation with  $Y$ . [See MULTIVARIATE ANALYSIS, articles on CORRELATION.]

**Plant selection.** In a plant-breeding program for improving the yield of a particular crop, a large number of potential new varieties become available in any year. These are tested in successive plantings, and the better-yielding types in each planting are selected to produce the seed for the next sowing, until finally a small fraction of the original number remain as possible replacements for the standard varieties in commercial use.

Finney (1966) and Curnow (1961) have carried out an extensive theoretical and numerical investigation of a fairly general type of selection procedure that is particularly applicable to plant

selection. The problem is to reduce in  $k$  stages an initial set of  $N$  candidates to a predetermined fraction,  $\pi$ , called the "selection intensity," using a fixed total expenditure of resources,  $A$ . At stage  $r$  ( $1 \leq r \leq k$ ), the candidates selected at stage  $r-1$  are tested, using resources  $A_r$ , to obtain for each a score,  $X_r$ , which estimates  $Y$  with a precision dependent upon  $A_r$ ; the fraction,  $P_r$  ( $0 < P_r \leq 1$ ), having the highest scores are selected, and the remainder are discarded. Stage 0 consists in selecting at random a fraction,  $P_0$ , of the initial set. The problem is to choose  $P_r$  and  $A_r$  to maximize the expected mean  $Y$ -value in the selected group, subject to  $A_1 + \dots + A_k = A$  and  $P_0 P_1 \dots P_k = \pi$ , where  $A$  and  $\pi$  are given. The authors found that approximately optimum results were obtained with  $A_r = A/k$ ,  $P_r = \sqrt[k]{\pi}$ , for  $r = 1, \dots, k$ , called the "symmetric" scheme, usually with  $P_0 = 1$ , although in some circumstances a value  $P_0 < 1$  effected further improvement. Three or four stages at most were sufficient.

In the context of plant selection, the  $N$  candidates are the new crop varieties produced in a particular year, and the resource expenditure,  $A$ , is the area of land available for testing, which must be divided into separate portions for varieties being tested for the first time, varieties selected on the basis of last year's tests to be tested in stage 2, and so on. For example, suppose  $N = 200$  varieties are started in a two-year program to select 8 to compare with the standard commercial types (thus,  $\pi = .04$ ). Then at the end of each year,  $\sqrt{.04} = \frac{1}{5}$  of the varieties should be selected—that is, 40 at the end of the year 1 and 8 at the end of year 2—with equal areas of land to be divided among the 200 varieties in year 1 and the 40 varieties in year 2.

**Drug screening.** In drug screening, the problem is to screen a large supply of chemical compounds by means of a biological test, usually in laboratory animals, in order to select for further testing the few that may possess the biological activity desired. Here  $Y$  stands for the unknown activity level of a compound (averaged over a conceptual population of animals), estimates being provided by the test results,  $X_1, X_2, \dots$ . The distribution of  $Y$  in the population of compounds available for screening will usually have a large peak at  $Y = 0$ , since most of the compounds do not possess the activity being sought unless a specific class of compounds chemically related to known active compounds is being screened. The number of compounds available for testing usually exceeds the capacity of the testing facilities; therefore, part of the problem in drug screening is to determine the optimum number,  $N$ , of compounds to screen in a given period of time.

The mean  $Y$  in the selected group does not have as much relevance in drug screening as it does in plant selection. Instead, a value,  $a$ , is usually specified such that a drug is of interest if its activity equals or exceeds level  $a$ . The screening procedure is then designed to maximize the number of compounds in the selected group having  $Y \geq a$ , usually subject to the requirement that the total number of compounds selected over a certain period of time is fixed. Davies (1958) and King (1963) have considered in detail this approach to the statistical design of drug-screening tests.

**Educational selection.** The consideration of selection procedures to allocate school children to different "streams" is necessarily much more complex than in the applications considered above. For one thing, there can be no question of rejection; the object, at least in principle, is to provide the education most suitable for each child. Furthermore, it can be expected that the characteristic  $Y$  of each child will be altered by the particular stream in which he may be placed. An admirable discussion of the problems was given by Finney (1962), who described, as an illustration of the methodology, a simplified mathematical model of the educational selection process then in operation in the British school system.

Finney considered university entrance as a two-stage selection process: the first stage is the separation of students at the age of 11+ into those who will receive a grammar school education and those who will go instead to a secondary school, and the second stage is university entrance. Denoting by  $X_1$  the composite score of all test results available at the first stage, by  $X_2$  the composite score at university entrance, and by  $Y$  the "suitability" of a student for university study as determined by his subsequent university grades, Finney considered  $Y$ ,  $X_1$ , and  $X_2$  to have a multivariate normal distribution with correlation coefficients estimated from available data. He studied the effects that varying the proportion of students admitted to grammar school, as well as the relative proportions admitted to universities from the two types of school, had on the average value of  $Y$  in the selected group and on the proportion of university entrants having  $Y \geq a$ . Interesting numerical results are presented, but the main feature of the paper is its demonstration of how the approach can bring about a clearer insight into the issues involved in a selection process.

#### The case where $N$ is small

Rather different methods of approach have been developed for selection when the number  $N$  is small

enough so that the individual values of  $Y$ , rather than their distribution, may be considered. The object of the selection procedure is expressed in terms of the  $Y$ 's; for example, the object may be to select the individual with the largest value of  $Y$ , to select the individuals with the  $t$  largest values, or to rank the  $N$  individuals in order according to their values of  $Y$ . These are special cases of the general goal of dividing the  $N$  individuals into  $c$  categories, containing respectively the  $n_c$  individuals with the highest values of  $Y$ , the  $n_{c-1}$  individuals with the next highest values, and so on, down to the  $n_1$  individuals with the lowest values of  $Y$ . Bechhofer (1954) developed expressions for the probability of a correct assignment of the  $N$  individuals to the  $c$  categories.

Selecting the "best" of  $N$  candidates. Of most frequent interest in practical applications is the selection of the best of several candidates, "best" being interpreted to mean the one with the largest  $Y$ . Estimates of the unknown  $Y$ 's are obtained from experimentation, and the problem usually is to decide how much experimentation needs to be done.

In a single-stage selection procedure, an experiment consisting of taking  $n$  observations for each candidate is performed, and the candidate with the highest observed mean is selected. The probability of a correct selection is the probability that the candidate with the highest  $Y$  will also have the highest observed mean value; this is a function not only of the number  $n$  of observations but also of the unknown configuration of values of  $Y$ .

Bechhofer (1954) recommended that the experimenter choose  $n$  so that the probability of a correct selection would exceed a specified value,  $P$ , whenever the best value of  $Y$  exceeded the others by at least a specified amount,  $d$ . His paper contains tables of the required value of  $n$ , calculated on the assumption that the  $Y$ 's are in the "least favorable" configuration, which in this case is the configuration where all the  $Y$ 's except the best one are equal and are less than the best one by the amount  $d$ .

Another approach to determining the value of  $n$  is based on striking an optimum balance between the cost of taking observations, which is a function of  $n$ , and the economic loss incurred if some candidate other than the best one is selected. The loss due to an incorrect selection is assumed to be a function, usually linear, of the difference between the largest  $Y$  and the selected  $Y$ . With the probability of selecting any particular candidate taken into account, an expected loss or risk function, which is a function of  $n$  and the unknown  $Y$ 's, is determined. Somerville (1954) showed how the minimax principle can be used to determine the



optimum sample size; this procedure is appropriate if no prior information is available about the unknown  $Y$ -values [see DECISION THEORY]. In the case where prior information is available, Dunnett (1960) showed how such information can be utilized in determining the sample size; he also made numerical comparisons between alternative procedures.

Gupta (1965) considered the situation in which the experimenter is willing to select a larger group than is actually needed and is concerned with guaranteeing a specified probability,  $P$ , that the best candidate is included in the selected group. In this way, the need to specify a minimum difference,  $d$ , as in Bechhofer's method, is avoided, but there is the drawback of not necessarily having a unique selection. Gupta investigated the effect of the sample size,  $n$ , and the configuration of the  $Y$ 's on the expected size of the selected group.

Sequential procedures have also been investigated; for example, Paulson (1964) considered a sequential method for dropping candidates from contention at each stage until only one remains, so as to achieve a specified probability,  $P$ , that the best one is selected whenever its value  $Y$  exceeds the others by at least  $d$ . [See SEQUENTIAL ANALYSIS.]

**A medical selection problem.** An interesting method (Colton 1963) for selecting the better of two medical treatments uses some of the principles discussed above but also contains ingenious points of difference. In the problem considered, there is a fixed number of patients to be treated. A clinical trial is performed on a portion of them, with equal numbers being given each treatment. On the basis of the trial, one of the two treatments is selected to treat the remainder of the patients. The problem is to determine how many patients to include in the trial in order to maximize the expected total number receiving the better drug. Sequential procedures for making the selection are also discussed.

## Tournaments

A tournament is a series of contests between pairs of  $N$  players (or teams) with the object either of selecting the best player or of ranking the players in order. It may be regarded as a selection procedure in which the experiments consist of paired comparisons between candidates. David (1959) studied some properties of two types of tournaments, the knockout and the round robin. Glenn (1960) compared the round robin and several variations of the knockout tournament for the case of four contestants; he found that a single knockout tournament with each contest being on a "best two

out of three" basis achieved the highest probability that the best player will win, but at the "expense" of requiring a higher average number of games.

## Other problems

There are many other interesting topics in screening and selection. One is *group screening*, in which a single test is performed on several candidates as a group to determine whether any of them possess the characteristic of interest. When an affirmative answer is obtained, further tests are performed to determine which ones possess the characteristic. One application of this procedure is in blood testing for the presence of some disease; a great saving in the number of tests necessary is accomplished by physically pooling several samples and making a single test. Another application is in factor screening in industrial research (see Sobel & Groll 1959; Watson 1961).

In practice, many screening problems are multivariate—that is, there is more than one trait,  $Y$ , of interest, and the traits are likely to be correlated. Sometimes the measurements on the several traits are reduced to a single variate by combination into a suitable index, perhaps with weights determined by the economic worth of each trait. Sometimes only one trait is dealt with at a time, the candidates considered for selection on the basis of trait  $Y$ , being those who have previously been selected on the basis of  $Y_1, \dots, Y_{r-1}$  in turn. Much work remains to be done to determine the best procedures for use in multivariate screening. (See Rao 1965 for a treatment of some of the mathematical problems.)

C. W. DUNNETT

[See also CLUSTERING; MULTIVARIATE ANALYSIS, article on CLASSIFICATION AND DISCRIMINATION; STATISTICAL ANALYSIS, SPECIAL PROBLEMS OF, article on OUTLIERS.]

## BIBLIOGRAPHY

- BECHHOFFER, R. E. 1954 A Single-sample Multiple Decision Procedure for Ranking Means of Normal Populations With Known Variances. *Annals of Mathematical Statistics* 25:16–39.
- COCHRAN, W. G. 1951 Improvement by Means of Selection. Pages 449–470 in *Berkeley Symposium on Mathematical Statistics and Probability, Second, Proceedings*. Berkeley: Univ. of California Press.
- COLTON, THEODORE 1963 A Model for Selecting One of Two Medical Treatments. *Journal of the American Statistical Association* 58:388–400.
- CURNOW, R. N. 1961 Optimal Programmes for Varietal Selection. *Journal of the Royal Statistical Society Series B* 23:282–318. → Contains eight pages of discussion.

- DAVID, H. A. 1959 Tournaments and Paired Comparisons. *Biometrika* 46:139-149.
- DAVIES, O. L. 1958 The Design of Screening Tests in the Pharmaceutical Industry. *International Statistical Institute, Bulletin* 36, no. 3:226-241.
- DUNNETT, C. W. 1960 On Selecting the Largest of  $k$  Normal Population Means. *Journal of the Royal Statistical Society Series B* 22:1-40. → Includes 10 pages of discussion.
- FEDERER, WALTER T. 1963 Procedures and Designs Useful for Screening Material in Selection and Allocation, With a Bibliography. *Biometrics* 19:553-587.
- FINNEY, D. J. 1962 The Statistical Evaluation of Educational Allocation and Selection. *Journal of the Royal Statistical Society Series A* 125:525-549. → Includes "Discussion on Dr. Finney's Paper," by T. Lewis et al.
- FINNEY, D. J. 1966 An Experimental Study of Certain Screening Processes. *Journal of the Royal Statistical Society Series B* 28:88-109.
- GLENN, W. A. 1960 A Comparison of the Effectiveness of Tournaments. *Biometrika* 47:253-262.
- GUPTA, SHANTI S. 1965 On Some Multiple Decision (Selection and Ranking) Rules. *Technometrics* 7:225-245.
- KING, E. P. 1963 A Statistical Design for Drug Screening. *Biometrics* 19:429-440.
- PAULSON, EDWARD 1964 A Sequential Procedure for Selecting the Population With the Largest Mean from  $k$  Normal Populations. *Annals of Mathematical Statistics* 35:174-180.
- RAO, C. R. 1965 Problems of Selection Involving Programming Techniques. Pages 29-51 in *IBM Scientific Computing Symposium on Statistics*, Yorktown Heights, N.Y., 1963 *Proceedings*. White Plains, N.Y.: IBM Data Processing Division.
- SOBEL, MILTON; and GROLL, PHYLLIS A. 1959 Group Testing to Eliminate Efficiently All Defectives in a Binomial Sample. *Bell System Technical Journal* 38:1179-1252.
- SOMERVILLE, PAUL N. 1954 Some Problems of Optimum Sampling. *Biometrika* 41:420-429.
- WATSON, G. S. 1961 A Study of the Group Screening Method. *Technometrics* 3:371-388.

## SEASHORE, CARL E.

Carl Emil Seashore (1866-1949), psychologist and leader in the development of American graduate education, was born in Morlunda, Sweden, where his family had been farmers for several generations. His parents migrated to Boone County, Iowa, in 1869, settling on an 80-acre farm in a Swedish community. Seashore's father acted as a lay Lutheran preacher and served as carpenter and cabinetmaker; he also had a considerable interest in music.

Seashore's early education, in Swedish, included the memorization of chapters of the Bible, a cherished feat in Sweden. At the age of 13 he went to live in the home of a Lutheran pastor "to improve his English, manners, and music." A year later he

became a church organist. He helped to run his father's farm for a few years and then attended the Lutheran college Gustavus Adolphus at a time when, according to his autobiography, "within an area of fifty square miles around our home only one man had gone to college . . ." (1930, p. 240). He received his A.B. in 1891.

Seashore went on to Yale to study philosophy with George T. Ladd, who introduced him both to the new branch of philosophy called experimental psychology and to Edward W. Scripture, a new faculty member not much older than Seashore. Seashore, nearing 26, was stimulated by working both with a mature scholar like Ladd and with Scripture, who after studying at Leipzig with Wundt had come to Yale to start a psychology laboratory. In Seashore's five years at Yale a three-fold foundation for his career as a psychologist was built. First, as Scripture's research assistant, he shared the problems of developing the laboratory course for psychology. In the laboratory he experimented in classical, physiological psychology through studies of visual accommodation time, perception of pressure and weight, and the laws determining illusions and hallucinations. He worked with Scripture on a spark chronometer and an audiometer. Second, in Ladd's seminar he undertook to defend the point of view that whereas evolution holds sway in all organic life, it does not apply to mental life. His conclusions, however, turned out to be the opposite of his initial hypotheses. Thus he developed a scholarly logic and a scientific Darwinian philosophy.

While at Yale Seashore lived with a Swedish family and attended the Lutheran church. Here the third element of the foundation for his career was developed. Although unwilling to preach sermons, he at times lectured the congregation on the relationship of philosophy and psychology to everyday life and developed a form of presentation understandable to an intelligent but untrained audience.

In 1895 he received his Ph.D. and spent that summer visiting psychology laboratories in France and Germany. He then returned to Yale as a fellow in psychology.

Two years later Seashore became assistant professor of philosophy at the State University of Iowa and thus an associate of George T. W. Patrick, professor of mental and moral philosophy and didactics at Iowa since 1887. Having taken work in psychology at Johns Hopkins with G. Stanley Hall, Patrick favored a strong development of the subject at Iowa and had set aside funds with which to build up a laboratory. Seashore developed several



important pieces of apparatus for use in a laboratory course and for research studies. This original equipment soon made the Iowa Psychology Laboratory prominent in the profession. Of special note was the equipment designed for the study of musical abilities.

In addition to his five papers on the problems of perception, published while he was at Yale, Seashore published papers in the University of Iowa Studies in Psychology, ten of them by the time he was promoted to professorial rank in 1902.

In 1905 he became chairman of the department of philosophy and psychology, succeeding Patrick. Under Seashore's chairmanship, the department increased the number of higher degrees awarded. This strong interest in promoting graduate work led to his appointment, in 1908, as dean of the graduate school. In addition, he served as chairman of a committee on the teaching of psychology appointed by the American Psychological Association and in 1911 as the association's president.

Seashore was a national leader in the field of graduate education as well as psychology. Under the auspices of the National Research Council, for example, he originated and carried out a project for gifted students, arguing that colleges and universities should devote as much attention to gifted students as to poor ones. He traveled widely both in his capacity as dean and in his efforts on behalf of the gifted student. Yet during this entire period he continued to do research and to publish.

He saw possibilities for the enrichment of teaching and research by developing the psychological aspects of several fields of learning and human welfare. A psychological clinic was established to supplement the services undertaken by the local psychiatric hospital. Collaboration with the speech department resulted in a speech clinic. Seashore initiated and supported an experimental approach to many problems associated with the Iowa School of Music, and a section for applied psychology of music was organized. Similar developments took place in graphic arts, plastic arts, and physical education.

Among Seashore's honors were several honorary degrees, including an sc.D. from Yale in 1935 and an Mus.D. from the Chicago Musical College in 1939. He was elected to the National Academy of Sciences and was a member of the Acoustical Society of America and the American Musicological Society, among others. He was also an honorary fellow of the British Psychological Society.

WALTER R. MILES

[For the historical context of Seashore's work, see the biography of HALL.]

#### WORKS BY SEASHORE

- 1893 On Monocular Accommodation Time. Yale University Psychological Laboratory, *Studies From the Yale Psychological Laboratory* 1:56-70.
- 1895 Measurements of Illusions and Hallucinations in Normal Life. Yale University Psychological Laboratory, *Studies From the Yale Psychological Laboratory* 3:1-67.
- 1896a Weber's Law in Illusions. Yale University Psychological Laboratory, *Studies From the Yale Psychological Laboratory* 4:62-68.
- 1896b Influence of the Rate of Change Upon the Perception of Differences in Pressure and Weight. Yale University Psychological Laboratory, *Studies From the Yale Psychological Laboratory* 4:27-61.
- 1897 A New Factor in Weber's Law. *Psychological Review* 4:522-524
- (1908) 1935 *Elementary Experiments in Psychology*. Rev. & enl. New York: Holt
- 1912 Measure of a Singer. *Science New Series* 35:201-212
- (1913) 1925 *Psychology in Daily Life*. New York: Appleton.
- 1919 *The Psychology of Musical Talent*. Boston and New York: Silver, Burdett and Co.
- 1923 *Introduction to Psychology*. New York: Macmillan.
- 1927 *Learning and Living in College: Psychology of Individual Differences Applied to the Organization and Pursuit of Higher Education*. Iowa University Series on Aims and Progress of Research, No. 21. Iowa City, Iowa: The University.
- 1930 *Autobiography*. Volume 1, pages 225-297 in *A History of Psychology in Autobiography*. Worcester, Mass.: Clark Univ. Press; Oxford Univ. Press.
- 1931 *The Graduate College in the State University of Iowa: A Survey of Thirty Years With Forecast*. Pages 9-83 in John W. Ashton (editor), *Trends in Graduate Work*. Iowa University Series on Aims and Progress of Research, No. 33. Iowa City, Iowa: The University.
- 1938a *A Preview to College and Life*. Iowa University Series on Aims and Progress of Research, No. 55. Iowa City: Univ. of Iowa Press.
- 1938b *Psychology of Music*. New York: McGraw-Hill.
- 1940 *The Junior College Movement*. New York: Holt.
- 1941 *Why We Love Music*. Philadelphia: Ditson.
- 1942 *Pioneering in Psychology*. Iowa University Series on Aims and Progress of Research, No. 70. Iowa City: Univ. of Iowa Press.
- 1947 *In Search of Beauty in Music: A Scientific Approach to Musical Esthetics*. New York: Ronald Press.

#### SUPPLEMENTARY BIBLIOGRAPHY

- BATHURST, J. E.; and SINCLAIR, R. D. 1928 A Complete Annotated Bibliography of the Writings of Carl Emil Seashore. *Psychological Monographs* 39, no. 2:3-22.
- MILES, WALTER R. 1956 Carl Emil Seashore: 1866-1949 Volume 29, pages 265-316 in *National Academy of Sciences, Biographical Memoirs*. New York: Columbia Univ. Press. → Includes a bibliography.

#### SEASONAL ADJUSTMENT OF TIME SERIES

See under TIME SERIES.

## SECHENOV, IVAN M.

Ivan Mikhailovich Sechenov (1829–1905), Russian physiologist, neurologist, and founder of objective physiological psychology, was born the son of a small landowner in the village of Tyoply Stan (now Sechenovo) in Simbirsk Province (now Arzamas Region). Taught first by private tutors, he mastered German and French in addition to Russian. At 14 he entered the St. Petersburg Military Engineering School. He served from 1848 to 1850 as a field engineer in a Kiev brigade. Then he resigned his military commission and from 1850 to 1856 studied medicine at the University of Moscow. He spent the next three and a half years doing post-graduate research and studying physiology, chemistry, and physics at Berlin, Heidelberg, Leipzig, and Vienna. Among his teachers were Johannes Müller, Emil DuBois-Reymond, Hermann von Helmholtz, Carl F. W. Ludwig, Robert W. Bunsen, and Heinrich Magnus.

Sechenov returned to Russia in 1860 to become at first adjunct and then full professor of physiology in the St. Petersburg Medico-Surgical Academy. Ten years later he resigned from that institution, giving as the reason, according to his *Autobiographical Notes* (1907), his distaste for its administrative policy and citing specifically the veto of his proposal to appoint Ilya I. Mechnikov to a vacant chair. (Mechnikov presumably was turned down because he was a Jew; later, in 1908, he won the Nobel prize.) For a while, Sechenov was without a professorial position and worked in the chemistry laboratory of his friend Dmitri Mendeleev, but he was soon appointed professor of physiology at the University of Odessa. He left that university in 1876 for the University of St. Petersburg and stayed there until 1888, when he resigned—alienated again by the academic atmosphere of St. Petersburg—and in 1889 moved finally to the University of Moscow.

In 1862 Sechenov took a year's leave to do experimental work in the laboratory of Claude Bernard at the Collège de France. There he discovered "central inhibition," namely, that spinal reflex movement may be "inhibited"—diminished or stopped—by the stimulation of higher-neural inhibitory centers. Further experimentation led to the important conception of higher-neural action as an interplay of excitation and inhibition. [See NERVOUS SYSTEM, article on STRUCTURE AND FUNCTION OF THE BRAIN.] This view was taken over wholly by Pavlov and more recently has won increasing acceptance as a result of direct microelectric neural

probing. Sechenov's report on inhibition was published in German and French (1863a).

Sechenov's physiological contributions soon became part of Western science. This is not true of his psychological essays—"Reflexes of the Brain" (1863b), "Who Must Investigate the Problems of Psychology, and How" (1873a), "Elements of Thought" (1878), and a number of others—which presaged the systems of psychology of Pavlov and Bekhterev and radically affected the world views not only of such distinguished men of science as Mendeleev, Mechnikov, Vladimir O. Kovalevskii, and Arkadii K. Timiriyazev but also of such widely known men of letters as Turgenyev, Gorki, and Tolstoi. Translated in part into French in 1884 and into English only in 1935, the radical Sechenov system of psychology was for years almost unknown outside Russia and had no influence whatever on Western and American thought in the field, from Wundt through James to Watson and the neobehaviorists. Only in 1950, in the second edition of Boring's *A History of Experimental Psychology* (1929), is Sechenov first mentioned in English in any significant sense. Recently, in *A History of Psychology*, by Esper (1964), a nine-page passage is devoted to his contributions, and it is asserted that Sechenov "wrote the first 'objective' psychology and became the first 'behaviorist' of modern times" (p. 324).

Briefly stated, Sechenov's psychological system rests on five interrelated theses:

(1) A consistent physical monism: Sechenov held that psychology will become a science only insofar as it studies the muscular and neural action of the psyche. He asserted that no "conceivable demarcation [from a scientific point of view] can be found between obvious somatic, i.e., bodily, nervous acts and unmistakable psychical phenomena" (1873b, p. 151).

(2) Physiological and psychical reactions are both considered to be reflex actions: in Sechenov's words, "All movements bearing the name of voluntary in physiology are reflex in a strict sense" ([1907] 1965, p. 109), and a thought "is the first two-thirds of a psychical reflex" ([1863b] 1935, p. 321).

(3) The reflex as the mechanism of association: "An association is," according to Sechenov, "an uninterrupted series of contacts of the end of every preceding reflex with the beginning of the following one" (*ibid.*, p. 312). The actual experience of an event and the memory of it are both represented by identical psychical reflexes but are evoked by different stimuli.



(4) The psychic as associative in genesis and central neural in mediation: perception and ideas emerge from associations of reflexes and their integration in the sensory sphere; both association and integration are mediated by the central nervous system.

(5) A radical environmentalism: the largest part of thoughts and ideas, 999 parts out of 1,000, as Sechenov would have it, derives from training, and only a minimal part is due to heredity.

Of these five theses, the first and last do, of course, have roots in earlier Western thought; the middle three are essentially original contributions by Sechenov.

GREGORY RAZRAN

[See also the biographies of BEKHTEREV and PAVLOV.]

#### WORKS BY SECHENOV

- 1863a *Physiologische Studien über die Hemmungsmechanismen für die Reflexthätigkeit des Rückenmarks im Gehirn des Froches*. Berlin: Hirschwald. → Also published as "Études physiologiques sur les centres modérateurs des mouvements réflexes dans le cerveau des grenouilles" in Volume 19 of the *Annales des sciences naturelles*.
- (1863b) 1935 *Reflexes of the Brain*. Pages 263–336 in Ivan M. Sechenov, *Selected Works*. Moscow: State Publishing House for Biological and Medical Literature. → First published as "Refleksy golovnogo mozga" in Volume 3 of the journal *Meditatsionnyi vestnik*.
- (1873a) 1935 *Who Must Investigate the Problems of Psychology, and How*. Pages 337–391 in Ivan M. Sechenov, *Selected Works*. Moscow: State Publishing House for Biological and Medical Literature. → First published as *Komu i kak razrabatyvat' psikhologiu?*
- 1873b *Psikhologicheskie etudy* (Psychological Studies). St. Petersburg: Sushchinskii.
- (1878) 1935 *Elements of Thought*. Pages 401–498 in Ivan M. Sechenov, *Selected Works*. Moscow: State Publishing House for Biological and Medical Literature. → First published as "Elementy mysli" in the journal *Vestnik evropy*.
- (1907) 1965 *Autobiographical Notes*. Washington: American Institute of Biological Sciences. → First published posthumously as *Avtobiograficheskie zapiski*.
- Izbrannye filosofskie i psikhologicheskie proizvedeniia* (Selected Physiological and Psychological Works). Moscow: Akademiia Nauk SSSR, 1947.
- Selected Works*. Moscow: State Publishing House for Biological and Medical Literature, 1935.

#### SUPPLEMENTARY BIBLIOGRAPHY

- BORING, EDWIN G. (1929) 1950 *A History of Experimental Psychology*. 2d ed. New York: Appleton. → See especially pages 635–636 on I. M. Sechenov.
- BUDILOVA, E. A. 1954 *Uchenie I. M. Sechenova ob oshchushchenii i myshlenii* (Sechenov's Views of Sensations and Thinking). Moscow: Akademiia Nauk SSSR.
- ESPER, ERVIN A. 1964 *A History of Psychology*. Philadelphia: Saunders.
- IAROSHEVSKII, M. G. 1961 *Problema determinizma v psikhofiziologii XIX veka* (The Problem of Determinism in the Psychology of the 19th Century). Dushanbe

(U.S.S.R.): Dushanbinskii Gosudarstvennyi Pedagogicheskii Institut.

- RAZRAN, GREGORY 1965 *Russian Physiologists' Psychology and American Experimental Psychology: A Historical and a Systematic Collation and a Look Into the Future*. *Psychological Bulletin* 63, no. 1:42–64.
- SHATERNIKOV, M. N. 1935 *The Life of I. M. Sechenov*. Pages ix–xxxvi in Ivan M. Sechenov, *Selected Works*. Moscow: State Publishing House for Biological and Medical Literature.

## SECTS AND CULTS

The term "sect" is used in the sociology of religion to designate a particular kind of religious group. This usage is both more precise and more technical than the use of the word in everyday speech. It is part of a typology of religious groupings which has been found useful in the study of religious movements and bodies. The term "cult" has also been given a special technical meaning as part of this typology, but it has remained less precisely defined, less useful, and less used in empirical research.

### Church and sect

In his study of the relationship between Christianity and the world, Ernst Troeltsch (1912) examined the tensions, problems, and dilemmas that confronted the Christian church as it attempted to come to terms with four aspects of classical civilization: family life, economic activity, politics and power, and intellectual endeavor. He found this history characterized by two major tendencies which exhibited themselves in varied forms over a long period of time. The first was the tendency to come to terms with secular society and culture, although often with considerable qualification, and generally to compromise with the world; the second, a highly significant rejection by a minority of the whole spirit of compromise and an opposition to important aspects of the secular culture and its institutions. These tendencies found organized expression in two basic and contrasting sociological types, which Troeltsch called the *church* and the *sect*.

The *church* represents the majority reaction and involves within its structure considerable variation of accommodation and compromise. It defines itself as the established expression of the relation between God and men, the institutional channel of divine grace, whose mission it is to enter the world in order to sanctify it. Thus, the church attempts to dominate the world with its values and is eventually dominated by the world to one extent or another. The church is characterized by what is vir-

tually membership on the basis of birth for the children of believers, although formally all are members through baptism. Thus, the practice of infant baptism is characteristic of churches, and churches become educational agencies. The church is the means for the administration of grace and exhibits the theological and sociological concomitants of this function: dogma and hierarchy. It is universal in its aspirations and addresses itself to the conversion of all. Consequently, its social structure is inclusive and often coincides with geographical or cultural entities or, as in the Middle Ages, with a whole civilization—Christendom.

The *sect*, as defined in the sociological literature since Troeltsch, represents a contrapuntal ideal type to the church form of social organization. It is a voluntary society of strict believers who live apart from the world in some way. Its foundation upon contracted or upon freely elected membership marks a sharp contrast with the ecclesiastical body of the church, as does its smallness of size and its spirit of austerity and asceticism. The sect expresses defiance of the world or withdrawal from it, a greater or lesser rejection of the legitimacy of the demands of the secular sphere. It emphasizes a conversion experience prior to joining. Sociological theory presents the church as the ideal type of religious body accommodating to the world and the sect as the ideal type of the protest group, protesting both the church's accommodation to the world and the world itself. The sect's refusal to compromise with secular values and institutions may find expression in either an active or a passive form. Hence there are two fundamental sect types: the militant oppositionist sect, which is active in its antagonism to the world, and the passive sect, which prefers withdrawal to militant defiance.

**Accommodation of sects.** H. Richard Niebuhr (1929), Liston Pope (1942), and others have delineated the sequential pattern in the course of which sects themselves are accommodated to the secular society and make their own compromise with the world. The birth of children to the sectaries, the better adjustment to conditions of life, including an increasing prosperity, and the passage of time itself—all contribute to a routinization of the sect into an established sociological entity which has accepted the social world in which it exists. Niebuhr observed that the sect, if defined rigorously in the terms presented here, cannot last beyond the founding generation. Pope's study would indicate that often considerable accommodation occurs in the lifetime of the founding generation. The routinized sect has been given the

name of *denomination*. Here once again, as in the use of the term *sect*, a word of common speech is given a more precise and technical designation. The typical picture presented in a number of sociological studies is that of the establishment of sects as protest groups against accommodation to the world and then their gradual acceptance of the world and their routinization as a reconciled part of it. This process is often associated with increasing wealth and respectability for the membership, in part at least the fruit of their ascetic and austere sectarian behavior.

**Established sects.** However, the work of J. Milton Yinger (1946) and Bryan R. Wilson (1959; 1961) has shown that not all sects, by any means, go through this sequence from sect to denomination and lose their earlier spirit of militancy and segregation. Some sects are successful in maintaining themselves over a long period of time in an established condition of opposition, or at least nonacceptance, with respect to secular society and its values. They become *established sects*, which, despite changes in their composition and milieu and the passing of the founding generation, retain a sectarian organization and an antagonistic or withdrawn posture in the face of the world. Of these established sects two types may be seen. There are those which secede from the world in a literal geographic sense and live in a territorial isolation in which they attempt to establish total communities after the model of their own values. The Amish, Hutterites, and others offer examples. There are also established sects such as the Jehovah's Witnesses and Christadelphians which remain within the general urban society and which nevertheless successfully maintain their opposition to it. Although such groups have not separated themselves geographically, they are separate in less palpable but not any less real ways, and they succeed in keeping their membership apart from genuine, intimate nonsectarian social participation.

**Sect ideology.** All sects display a considerable degree of totalism in dominating the lives of their membership. Ideological domination is usually supplemented and supported on the social level by measures which set the group apart, such as endogamy, limitations on the forms of participation with outsiders, refusal to take part in significant common societal activities (military service, saluting the flag, or medical practice), peculiar habits of eating and abstinence, and with some groups, even peculiarities of dress. Related to these social forms of segregation is the notion of the sect members as comprising the "elect," some kind of religious elite.



### Social bases of sectarianism

Sects are opposition groups, and they arise in opposition to the accommodation of churches or developing denominations, in rejection of some other aspects of their milieu, or in some combination of the two. Troeltsch has shown that the sect form asserted itself early in the Middle Ages. It is to be seen in the period of the Gregorian agitation (c. 1080) when the sectarianism of the Albigensians spread in Italy and France. This movement had complex social and religious sources. It was greatly affected by the reform efforts and struggles of Pope Gregory VII; it expressed the opposition of the devout laity to what they considered immorality and simony in the church; and it also represented the aggressive reaction of new urban classes against the established order in both church and city. This correspondence and interpenetration of religious and social interests has often been found associated with the origin and formation of sects. It has often been observed in the sociological literature that the sect is a lower-class protest phenomenon.

The conditions of life of different social strata influence the psychological make-up and need dispositions of their members. Consequently, social classes and strata develop different religious needs and sensibilities. Niebuhr stated that the religion of the disinherited may be observed in the rise of many sects and that Christianity was at first the religion of those who had little stake in the civilization of their time. Troeltsch concluded that all really creative religious movements are the work of lower strata. Niebuhr stressed the importance of economic success in the transformation of protesting sects into denominations and pointed to the fact that the churches of the poor sooner or later become churches of the middle class.

**Functions of the sect.** The sect exhibits complex functions in society. It often offers an outlet for strains and frustrations incumbent upon lower-class status and for the condition of being socially and economically disinherited. In allowing catharsis, it at the same time provides a meaningful community, together with a set of values that promotes a personal reorganization of the members' lives and often their eventual reincorporation into the general society. Not only may the sect reconcile the disinherited to their situation through the various compensations of this-worldly community and other-worldly expectations, but it may also bring new meaning to them in its reinterpretation of their life experience. In doing this it may socialize its members in virtues which lead to economic and

worldly success. Moreover, the sect, with its close community of human beings and its new values which give meaning to life, offers a way out of anomie to many who have been disorganized in the impersonal milieu of the modern city. When the founding generation passes away, the established sect continues to perform similar functions for individuals who are attracted to it and provides for its born members the setting for acting out their established values. Sects may take on a number of new functions when their social composition and their specific social situation change over time.

When established organizational conditions offer insufficient expression to the religious needs of people or when established institutions fail to meet needs of particular strata and groups at all, it is easy for charismatic leaders to arise and organize a following. Such developments issue in movements of protest of a marked sectarian character. The charismatic leader as a rallying focus and an active initiator plays a strategic role in the origin of sects and often impresses his own self-interpretation upon the group as the model for its behavior and beliefs. The accommodation and routinization of churches and the development of sects into denominations is often the occasion for schism, which is an important source of sectarian movements. Moreover, conditions and social change within the general society, altered economic status for particular groups, urbanization, increased mobility—geographic and psychological—and other phenomena associated with industrialization all contribute to the rise of sects.

The sect as a sociological ideal type is therefore to be understood as the embodiment and expression of rejection of some significant aspect of secular life. It represents a protest against compromise with the society and its values and the institutional development of the church itself as an aspect of this accommodation. It is charismatic, lay, egalitarian, and voluntaristic religion in contrast to the established, professional, hierarchical, and ascribed religion of the church. In this typology the sect represents an ideal type: empirical reality and specific historical development present a greater variety than does the typology itself.

### Allied types of religious protest

Many protest movements display sectarian characteristics but to different degrees and often in somewhat different respects. Most of the important protest movements in Christianity, while highly influenced by sectlike elements, endeavored to achieve organizational forms which also involved many of the characteristics of the church. Thus the

Reformed churches of the Protestant Reformation vary along a complex continuum from Anglicanism with its episcopate and quite ecclesiastical structure, at one end, to the sectlike organizations of the Baptists, at the other, with interesting combinations of church and sect attributes characterizing the in-between groups, for example, the Churches of the Standing Order in colonial Massachusetts. Joachim Wach (1944) has called a number of them *independent groups* and has pointed out that they vary in form from churchlike hierarchical structures to egalitarian covenants of laymen.

However, not all protest is secessionist in intention, nor does protest necessarily issue in separate organization outside the established bodies. Monasticism and the later religious orders offer an outstanding example of protest groups which remain within the older ecclesiastical body. Monasticism exhibits a number of sectarian qualities: it establishes a separate community, practices austerity and asceticism, and employs segregating rules and peculiarities of dress. Like the geographically isolated sects, it creates its own distinct community but remains dependent upon the larger body for replacement of personnel. In its origin Christian monasticism was both a protest against the accommodation of the church and a rejection of the world. Its relationship to the sacramental church was ambiguous, and it could have become a secessionist movement. But in the rule of Basil in the East and of Benedict in the West, it was reintegrated formally and solidly into the structure of the church. Here it continued to play a role of witness and to advocate reform. Moreover, it placed its enormous energy at the disposal of the church for missionary and other activity. In the High Middle Ages, the Franciscan movement represented a similar tendency. It was contained within the church at first by the personal character of its founder. Later on, its integration into the church was the cause of a great struggle in which both schism and heresy as well as reintegration of the order into the church resulted. Moreover, the routinization process from sect to denomination is also found in the history of religious orders. Such routinization is often the cause of schism and divisions and the rise of reforming leaders of the charismatic type.

**The Mormons.** A religious body of a marked sectlike character which seeks geographical isolation may, when circumstances are propitious, develop into an entity resembling an ethnic group or even a nation. The Mormons, a sectlike group choosing to imitate the Biblical model of Israel, found themselves in circumstances where such recapitulation took on realistic significance. Perse-

cuted and driven from their settlements, achieving victories and suffering defeats, the Mormons built up in a decade and a half a folk tradition and mentality of their own. In moving to the West they found a vast unoccupied expanse of land upon which they could expand their vision of an earthly kingdom of God to imperial dimensions. As a result the semiecclesiastical organization which developed was at the same time the organized core of a Mormon people held together by kinship ties, common beliefs and values, a common history of achievement and suffering, and a common homeland. The Mormon "Zionism" of the nineteenth century had led to a development from near sect to near nation. When the Mormons applied for admission of their state of Deseret to the federal union, they attempted to find a political form for their achievement which stopped just short of nationhood; and in times of stress and conflict, frankly separatist sentiment was widespread (O'Dea 1954). Churches have also become the core of ethnic groups, as under the Turkish millet system in the Middle East, which granted a degree of political autonomy to religious communities.

### Typological study of sects

Wilson has shown that it is possible to distinguish types of sects on the basis of their ideological orientations. He does this within the context of Protestantism on the basis of the sect's self-definition, specifically its conception of its calling and mission. He distinguishes first the *conversionist sect*, which seeks to convert others and thereby to change the world; second, the *adventist sect*, which expects drastic divine intervention and awaits a new dispensation; third, the *introversionist sect*, which is pietistic in its orientation, withdrawing from the world to cultivate its inner spirituality; and the *gnostic sect*, which offers some special esoteric religious knowledge. Such sects will experience the effects of routinization differently and will also exhibit different structural tendencies to some degree (Wilson 1959; 1961).

Moreover, since the terms *church* and *sect* are ideal-typical constructions, what is observed in real life situations only approximates the specifications of the theoretical definitions. Such ideal-typical concepts have an analogical character and are most useful for observation, analysis, and interpretation when utilized with flexibility. This analogical character of the Troeltschian concepts is best seen in the behavior of churches when placed in circumstances which elicit sectlike behavior from them. The Roman Catholic church in the United States in the nineteenth century found itself a minority



religion, largely lower class in character, constituted in its vast majority by ethnic groups of recent immigrant origin, and therefore of lower prestige in the general American society. Moreover, the value system of American society was largely derived from Protestantism, and the various forms of Protestantism constituted something like an unofficially established national religion. The Roman Catholic church responded by separating itself from the surrounding Protestant world in a wide range of activities and by constructing its own institutional contexts for education from the primary grades through the university, for social welfare work, for hospitals and other institutions for aid, and for sports and entertainment. Moreover, the mentality of American Catholics took on a number of sectlike attributes, such as apartness and defensiveness, rigorism in morality, and militancy in religious identification. While this situation was in part conditioned by the defensive character of post-Tridentine Catholicism in Europe and by the Irish background of so many American Catholics, there is no question of the importance of American conditions in bringing about a sectlike result. It is significant in this connection that the first heresy conflict in many decades to be seen in American Catholicism concerned a student center at Harvard University which displayed a militant sectarian response to secularism in the intellectual sphere and against the church's accommodation to the secular world. This group ended in heresy and excommunication (O'Dea 1961).

What has evolved from the time of Troeltsch is a typology of religious groups which has proven its utility in description and analysis in the sociological study of religion. It may be summarized briefly as follows: The *church* is the embodiment of institutional religion and accommodation to the world. It gives rise to protest groups and movements. These may vary from *reformed churches* to *independent groups* to *sects*, or they may give rise to groups which remain within the older body, affecting and reforming it in various ways. Sects may be either actively in opposition or passive and withdrawing with respect to the world. They may be geographically isolated, or they may exist within the general society, practicing forms of social separatism. Sects may go through a process of social mobility and routinization and develop into *denominations*, accepting in one way or another the secular society and its values. The sect may also institutionalize its oppositionist character and become an *established sect*. Under propitious circumstances a sect or near sect can develop into a

new ethnic or quasi-ethnic entity; it can become a *people*.

While this typology has been useful and represents the contributions and insights of a number of sociologists and scholars, it remains unsatisfactory. Like all ideal type concepts, it is unwieldy to use in analysis and possesses severe limitations with respect to refinement or adaptation to mathematical use. What is necessary is to analyze these ideal-typical constructs and to state their components in terms of factors or variables. This would mean replacing the global concept with a number of dilemmas or choice points which give rise to one kind of organizational tendency rather than another or, in a like manner, breaking up the ideal types into tendencies which vary from group to group along a series of continua. A similar kind of analysis was made by Talcott Parsons with respect to the ideal type conceptions of Tönnies (*Gemeinschaft and Gesellschaft*), which he broke down into five pattern variables. Such factors or variables would make it possible to compare groups with respect to a number of characteristics instead of the present, more cumbersome process of using ideal types as an analytical model.

**The cult.** To this already complex typology, von Wiese and Howard Becker (1932) have added the *cult*. While this term has been less clearly defined, it has been most generally used in the scholarly literature to designate a more loosely organized and more individualistic group than those already discussed (this usage must be distinguished from the scholarly use of the word *cult* to refer to the ritual act of worship). Based on individual concerns and experiences, the cult is often transient, its membership often highly fluctuating. Belonging to the cult often does not involve an acceptance of common discipline and need not necessarily preclude membership in other kinds of religious groups. Both theosophy and New Thought have been labeled as cults. Wilson (1961) considers the cult a gnostic sect and, in his study of Christian Science in Great Britain, shows a group that combines aspects of bureaucratized organization with characteristics of both sect and cult.

What is involved in this type of group may be best understood by returning to Troeltsch's original treatment, with which this article began. American sociologists have tended to make only a partial use of the paradigm of analysis introduced by Troeltsch and to develop only the implications of the church-sect dichotomy.

Troeltsch, however, emphasized another kind of religious reaction to accommodation to the world

and routinization in the development of religious forms of expression. He speaks of *mysticism*, which is found particularly when the formalization of worship and doctrine makes individual religious experience difficult and unfruitful within the established forms. Just as the sect tends to be a lower-class phenomenon (the religious form of those without a stake in the social system), so mysticism is characteristic of the educated classes. It has been an element enriching the life of the established religious bodies, as it has also been an expression of protest against them. It is of great significance in the development of the religious life of the Catholic church and was a tremendously important element in the Reformation and in post-Reformation Protestantism. When it does affect the lower classes and when found in religious movements of the poor, it often involves emotional excesses and a taste for heterodox novelties. The religious experience of worship and a relationship to the Deity, the gnostic experience involving secret knowledge and skill not available to all, and the mystical effort to achieve a personal relationship with God outside the established forms of worship and even of language are all obviously related. Consequently, some element of mysticism may be found in varied versions of the religious experience, although religious traditions placing their emphasis upon law often discourage mysticism, evidently fearing its antinomian possibilities.

**Sects in non-Christian cultures.** Most of the work on sects in the sociological literature concerns Christian groups. It is obvious, however, that the typology refers to aspects and characteristics of religious groups and movements to be found outside the Christian tradition. Wach has pointed out that Zoroastrianism and Mahāyāna Buddhism have produced ecclesiastical bodies which fit the Troeltschian definition of the church in its main outlines and that groups in Islam and Confucianism have evolved semiecclesiastical bodies. Monasticism has been developed in a number of quite different religious and cultural traditions. Sects which fit many of the characteristics we have presented above may also be found in the other world religions. For example, in Islam the rise of Wahhabism represents an active oppositionist sect, while Bahai arose as a sect of the passive withdrawing character.

**Sectarianism in politics.** Finally, it should be noted that the typology given here is to some extent applicable to organizations other than those of a religious character. Groups with situationally transcendent ideals seem to display similar organ-

izational types. This may be seen in political parties with some degree of utopianism in their programmatic aspirations. For example, social democratic parties in Europe were formed upon situationally transcendent ideals and in opposition to existing conditions. A degree of success both for membership and officialdom led to a process of routinization and accommodation analogous to that experienced by a church in the religious sphere (Michels 1911). As a consequence, protest movements developed, such as the Spartacus group in Germany, or even the Third International. Thus, the basic variables involved in this typology would appear to be found whenever organized expression of interests based upon situationally transcendent aspirations occurs.

THOMAS F. O'DEA

[Directly related are the entries MILLENARISM; RELIGIOUS ORGANIZATION. Other relevant material may be found in COLLECTIVE BEHAVIOR; MASS PHENOMENA; SOCIAL MOVEMENTS; and in the biographies of MANNHEIM; TROELTSCH; WEBER, MAX.]

#### BIBLIOGRAPHY

- BERGER, PETER L. 1954 *The Sociological Study of Sectarianism. Social Research* 21:467-487.
- BESTOR, ARTHUR E. JR. 1950 *Backwoods Utopias: The Sectarian and Owenite Phases of Communitarian Socialism in America: 1663-1829*. Philadelphia: Univ. of Pennsylvania Press.
- CLARK, ELMER T. (1937) 1949 *The Small Sects in America*. Rev. ed. Nashville, Tenn.: Abingdon-Cokesbury.
- COHN, NORMAN (1957) 1961 *The Pursuit of the Millennium: Revolutionary Messianism in Medieval and Reformation Europe and Its Bearing on Modern Totalitarian Movements*. 2d ed. New York: Harper.
- FRANCIS, E. K. 1950 *Toward a Typology of Religious Orders. American Journal of Sociology* 55:437-449.
- FREUND, WILLIAM H. C. 1952 *The Donatist Church: A Movement of Protest in Roman North Africa*. Oxford: Clarendon Press.
- JOHNSON, BENTON 1957 *A Critical Appraisal of the Church-Sect Typology. American Sociological Review* 22:88-92.
- JOHNSON, BENTON 1961 *Do Holiness Sects Socialize in Dominant Values? Social Forces* 39:309-316.
- MANN, WILLIAM E. 1955 *Sect, Cult and Church in Alberta*. Univ. of Toronto Press.
- MANNHEIM, KARL (1929-1931) 1954 *Ideology and Utopia: An Introduction to the Sociology of Knowledge*. New York: Harcourt; London: Routledge. → A paperback edition was published in 1955 by Harcourt. Part 1 was written for this volume. Parts 2-4 are a translation of *Ideologie und Utopie* (1929); Part 5 is a translation of the article "Wissenssoziologie" (1931).
- MICHEL, ROBERT (1911) 1959 *Political Parties: A Sociological Study of the Oligarchical Tendencies of Modern Democracy*. New York: Dover. → First published as *Zur Soziologie des Parteiwesens in der modernen Demokratie*.



- MUENDER, WALTER G. 1945 From Sect to Church. *Christendom* 10, no. 4:450-462.
- NIEBUHR, H. RICHARD (1929) 1954 *The Social Sources of Denominationalism*. Hamden, Conn.: Shoe String Press. → A paperback edition was published in 1957 by World.
- O'DEA, THOMAS F. 1954 Mormonism and the Avoidance of Sectarian Stagnation: A Study of Church, Sect, and Incipient Nationality. *American Journal of Sociology* 60:285-293.
- O'DEA, THOMAS F. 1961 Catholic Sectarianism: A Sociological Analysis of the So-called Boston Heresy Case. *Review of Religious Research* 3:49-63.
- POBLETE, RENATO; and O'DEA, THOMAS F. 1960 Anomie and the "Quest for Community": The Formation of Sects Among the Puerto Ricans of New York. *American Catholic Sociological Review* 21:18-36.
- POPE, LISTON 1942 *Millhands and Preachers: A Study of Gastonia*. New Haven: Yale Univ. Press; Oxford Univ. Press.
- TROELTSCH, ERNST (1912) 1931 *The Social Teaching of the Christian Churches*. 2 vols. New York: Macmillan. → First published as *Die Soziallehren der christlichen Kirchen und Gruppen*. A paperback edition was published in 1960 by Harper.
- WACH, JOACHIM (1944) 1951 *Sociology of Religion*. Univ. of Chicago Press.
- WEBER, MAX (1906) 1946 *The Protestant Sects and the Spirit of Capitalism*. Pages 302-322 in Max Weber, *From Max Weber: Essays in Sociology*. New York: Oxford Univ. Press. → First published in German.
- WEBER, MAX (1915) 1946 Religious Rejections of the World and Their Directions. Pages 323-359 in Max Weber, *From Max Weber: Essays in Sociology*. New York: Oxford Univ. Press. → First published in German.
- WHITLEY, OLIVER R. 1955 The Sect-to-Denomination Process in an American Religious Movement: The Disciples of Christ. *Southwest Social Science Quarterly* 36:275-281.
- WIESE, LEOPOLD VON (1924-1929) 1932 *Systematic Sociology: On the Basis of the "Beziehungslehre and Gebildelehre" of Leopold von Wiese*. Adapted and amplified by Howard Becker. New York: Wiley. → First published as *System der allgemeine Soziologie*.
- WILSON, BRYAN R. 1959 An Analysis of Sect Development. *American Sociological Review* 24:3-15.
- WILSON, BRYAN R. 1961 *Sects and Society: A Sociological Study of the Elgin Tabernacle, Christian Science, and Christadelphians*. Berkeley: Univ. of California Press.
- YINGER, J. MILTON (1946) 1961 *Religion in the Struggle for Power: A Study in the Sociology of Religion*. Sociological Series, No. 3. Durham, N.C.: Duke Univ. Press.

## SECURITIES MARKETS

Securities markets exist in more than sixty countries and on all continents. They have grown in number and importance with the increase in the number and size of corporations, the volume of government debt instruments, and the diffusion of ownership of corporate and government securities.

Securities markets are organizations of security

dealers and brokers whose operations reduce the cost of transferring ownership of government and corporate bonds and stocks and increase the liquidity of these assets. The reduction in transaction costs is achieved by the specialization of these organizations in transmitting information relevant for the decisions of potential buyers and sellers, by the provision of standardized contracts with respect to their purchase, sale, and financing, and frequently by furnishing specialized facilities for the execution of security transactions. Securities markets—especially organized exchanges—increase the liquidity of securities by providing almost continuous information regarding market prices, maximum bids, and minimum offers. Liquidity is increased in the sense that owners of securities can have greater knowledge of the current market value of their securities. This reduces the risk of incorrectly assessing the value of one's securities. This increase in liquidity is important not only for holders of existing securities but also for issuers of new securities who are raising capital for their enterprises.

Securities markets, especially those operating through major exchanges, have some of the characteristics of perfectly competitive markets. The things that are bought and sold—shares of common stock of the General Motors Corporation, for example—are perfect substitutes for each other, and the dissemination of information on market prices is accurate, extremely rapid, and very widespread. Information relevant for appraising the prospects of a corporation is also widely available. Further, there are many buyers and sellers, most of whom buy and sell in quantities insufficient to have a significant impact on the price. Although large institutional investors can temporarily raise the price of shares of even very large corporations by placing large "market" orders (and conversely depress the price of shares by selling), this impact can be greatly diminished by spreading the orders over a period of time. The liquidity of even very large holdings of common stock was dramatically illustrated in recent years by the sale of millions of shares of Ford Motor Company stock by the Ford Foundation with no apparent large impact on prices.

Security prices are highly sensitive to information bearing on the prospective profitability of the corporations whose shares are involved. The opportunities to profit from changes in the prices of securities have attracted the attention and the money of millions of investors and speculators and have generated many points of view about ways to predict future security prices. The numerous points of view fall into two general categories: (1) efforts

to predict the profitability of firms by forecasts of general business conditions, by forecasts of costs, revenues, and profits for particular industries, and by analyses of the competitive position of particular firms, and (2) efforts to predict future prices by an analysis of the historical course of prices themselves. This latter point of view has been subjected to considerable analysis by academic economists and statisticians (Cootner 1964), and the predominant view among academic economists is that historical analyses of prices do not enhance one's ability to predict prices. In the short run, prices are believed by many to follow a "random walk." [See MARKOV CHAINS.] This academic view is at variance with the views of many members of the financial community.

The view that prices follow a random walk is consistent with a market in which numerous pieces of relevant information bearing on future prospects are frequently generated and widely and rapidly disseminated. The pieces of information themselves, however, must be causally unrelated to the historical course of security prices. Another essential condition is that prices adjust "instantaneously" to new information, since gradual adjustment would create trends.

**Organized exchanges.** The most important securities markets in terms of the value of securities traded are organized exchanges. Such exchanges differ from less formal markets in that they centralize communication and trading, impose rules for the admission of persons or firms privileged to trade in securities, impose rules with respect to the securities that can be traded, and closely regulate the procedures used.

By far the largest securities market in the world in terms of the dollar volume of securities traded is the New York Stock Exchange. This exchange, started in 1792 as an informal organization of 24 merchants and auctioneers dealing in securities, has grown throughout most of its history. At the end of 1966 there were 1,366 members of the exchange, most of them associated with approximately 660 organizations, including single proprietorships, partnerships, and corporations. About 1,200 companies with approximately 1,600 different issues of stock were listed, and the more than 10,000 million shares listed had an aggregate value of more than \$495,000 million. In addition, about 500 companies and government agencies with approximately 1,200 different issues of bonds were listed, and these bonds had an aggregate market value of approximately \$120,000 million. At the end of 1966 there were more than 250 listings of foreign securities, including both stocks and bonds,

having an aggregate value of more than \$9,000 million.

The value of securities listed on the New York Stock Exchange at the end of 1966 was more than four times as great as the value of securities listed on the world's second largest exchange, The Stock Exchange (of London), which trades a larger number of issues and has more than twice as many members. The New York Stock Exchange accounts for more than 90 per cent of the dollar volume of trading on the 18 organized exchanges in the United States, and this proportion has been maintained for many years (U.S. Securities and Exchange Commission 1963a).

The number of members of the New York Stock Exchange has been fixed in recent years. Memberships, called "seats," are bought and sold. The price of a seat in 1966 was approximately \$250,000. In 1875 seats sold for as little as \$4,250, and in 1929 they reached an all-time high of \$625,000. To be eligible for membership, persons must meet standards established by the exchange with respect to age, citizenship, character, and knowledge of the securities business.

About one-half of the members of the New York Stock Exchange belong to organizations whose primary purpose is to act as brokers on behalf of those wishing to buy or sell listed securities. These members are paid according to a fixed schedule of brokerage fees averaging roughly one per cent of the value of the security bought or sold, although the fees vary somewhat in a prescribed manner according to the number of shares and the aggregate value of shares or according to whether the securities traded are stocks or bonds.

About one-fourth of the members are "specialists." They are important in the operation of the New York Stock Exchange and other American exchanges, but other devices are used outside the United States, except in Japan, to perform the specialist's function. The specialist receives information from brokers regarding offers to sell and to buy at various prices. Each specialist concentrates his activities in a small group of securities, and brokers wishing to transact business on behalf of the public typically deal with one or another of the specialists handling the stock in question, although brokers may deal directly with other brokers. The specialist secures for the seller the highest available bid and for the bidder the lowest available offer. When there is a wide discrepancy between the highest bid and the lowest offer, the specialist has the responsibility, neither fully enforceable nor precisely defined, to use his own capital to maintain a "reasonably continuous" mar-



ket in such a stock. The specialist is also permitted to buy and sell securities for the account of others, thus acting as a brokers' broker. Regulations of the Securities and Exchange Commission (SEC) require that the specialist must execute orders on behalf of the public prior to executing them for his own account.

Part of the justification given for the existence of the system of specialists is that their trading on their own behalf increases liquidity and reduces the likelihood of large changes in prices between successive transactions. A detailed investigation of the impact of specialists on prices during the period immediately following the assassination of President Kennedy in 1963 revealed that specialists vary widely in their trading practices and in their impact on prices (U.S. Securities and Exchange Commission 1963b). A substantial number of specialists traded so as to diminish volatility, and a substantial number so as to increase it.

Approximately one-fourth of the members of the New York Stock Exchange are either odd-lot dealers or "floor brokers." The former buy and sell blocks of stock of less than 100 shares (odd lots) from brokers representing the public. The latter assist brokers from the commission houses when the volume of business exceeds their capacity to handle it.

Most transactions involve a small enough volume of stock so that the ordinary market mechanism can be used without having a large impact on prices. Transactions, however, which involve very large numbers of shares and dollars are often handled through specialized procedures designed to lessen this impact. The most important type of block distribution is a "secondary distribution" (the sale by an underwriter or group of underwriters of a large volume of stock *not* for the benefit of the corporation whose shares are being sold), but "exchange distributions" (within brokerage houses) and other specialized means of dealing with large volumes of stock are growing in importance. Although the details of these various specialized methods differ, they have in common an effort to accumulate orders to buy from large numbers of investors over a period of time in order to match the large volume of stock that is being offered by a single seller. For instance, in mid-1964, 250,000 shares of a stock selling for \$33 were offered by a single seller. The stock was distributed within one day through the facilities of a single brokerage organization whose customers bought all the stock. The individual purchases were bunched and executed as a package on the floor of the New York Stock Exchange. They were represented as a single

transaction through the national and international communications system that virtually instantaneously transmits information on stock prices. This particular method of distributing stock is known as an "exchange distribution."

The number of shares traded as a percentage of the number of shares listed has been declining secularly. Around 1900, the number of shares traded in a year was approximately 200 per cent of the average number of shares listed in that year, and as recently as 1929 the number of shares traded during a year exceeded the average number listed. Since World War II the number of shares traded in any year has never exceeded 25 per cent of those listed, and in recent years only about 15 per cent of listed shares have been traded in any year.

In order for the shares of a company to achieve initial listing on the New York Stock Exchange, the company must demonstrate earning power, after charges and taxes, of \$1.2 million annually; must have net tangible assets of \$10 million, or more than \$10 million in market value for publicly held shares; and must have at least 600,000 shares publicly held and not fewer than 1,500 owners, each of whom must own at least 100 shares. Similar though less stringent criteria are applied for delisting a stock.

For numerous periods between January 1926 and December 1960 average rates of return to investors investing equal amounts of money in each of the common stocks listed on the New York Stock Exchange were generally higher than earnings in alternative investments, such as government and corporate bonds, savings accounts, and mortgages (Fisher & Lorie 1964). For the entire 35-year period, the average annual rate of return (compounded) to a tax-exempt investor, with reinvestment of dividends, was 9.0 per cent. Without reinvestment of dividends the rate was 6.9 per cent. During the decade of the 1950s, the average annual rates with and without reinvestment were about 15 per cent. (Rates of return with reinvestment may be higher, lower, or the same as without reinvestment. The rate of return is based on the relationship between a change in the value of assets and the volume of assets initially invested. Reinvestment increases both the change in value and the amount invested, and the effect on the rate of return depends on the relative magnitude of these increases.)

Data are also available on the variability of rates of return (Fisher 1965). Considering all possible purchases and sales, at monthly intervals, of all common stocks on the New York Stock Exchange

for the period 1926–1960, it was found that slightly over three-fourths of the transactions were profitable and over half provided rates of return in excess of 9 per cent. Of course, this past experience is not a certain guide to the future.

The Stock Exchange (of London) lists more than 9,000 different securities that had at the end of 1963 a market value of approximately \$125,000 million, about a third of which was represented by debt instruments of the government of Great Britain, Commonwealth governments, local councils, and nationalized industries. The Stock Exchange had about 3,400 members in mid-1964. The relative importance of foreign shares is greater for this stock exchange than for the New York Stock Exchange. The major difference in the method of operation of The Stock Exchange—and almost all exchanges outside the United States—as compared with the New York Stock Exchange is that brokers who buy and sell securities on behalf of the public deal with jobbers who own securities and act as wholesalers. These jobbers, who do not deal directly with the public, serve much the same purpose as the specialist on the New York Stock Exchange, but the specialist in the United States typically has a dual role—broker and dealer—while the jobber in London operates solely as a dealer adding to or depleting his inventory of the securities in which he trades. There are about 280 firms of brokers and about 65 firms of jobbers operating on The Stock Exchange.

A jobber may deal in any listed security, but most specialize in certain groups of shares, or “markets.” The chief markets are as follows: gilt-edged (government issues and other securities grouped with them); banks and insurance; shipping; foreign government bonds; American and Canadian shares; breweries; commercial and industrial; iron, coal, and steel; financial, land, and property; investment trusts; rubber and tea plantations; oil; South African mines; Rhodesian, Canadian, Australian, and miscellaneous mines; West African mines; cables and transportation.

Rates of return on British shares between 1919 and 1963, averaged by 11-year periods, were substantially higher in most of these periods—excluding wartime—than rates on consols—government bonds with no maturity date (Merrett & Sykes 1963). For the entire period, the rates on shares were 8.0 per cent in money terms and 5.8 per cent in real terms, while the comparable rates for consols were 1.4 per cent and –1.4 per cent.

There are other important exchanges in the world, but none approaches in importance either the New York Stock Exchange or The Stock Ex-

change (of London). Among the other leading exchanges are the Stock Exchange of Paris, the Brussels Stock Exchange Commission, the Zurich Stock Exchange (the leading Continental market), the Dusseldorf and Frankfurt stock exchanges in the Federal Republic of Germany, the Milan Stock Exchange, the American Stock Exchange (United States), the Toronto and Montreal exchanges in Canada, and the Melbourne Stock Exchange. There are many regional exchanges in the countries whose leading exchanges are listed above, and there are many exchanges in other countries as well. The largest of these exchanges is not more than about one-fifth the size of The Stock Exchange (of London) or one-twentieth the size of the New York Stock Exchange, and most are much smaller. The mechanisms of these various exchanges differ, but the principles underlying their operation are very much like either The Stock Exchange or the New York Stock Exchange. Information about many of these exchanges can be found in *Spray* (1964).

**Over-the-counter markets.** Unlike the organized exchanges, the over-the-counter market is informal and permits unlimited right of entry by securities and virtually free access of persons into the business of dealing in unlisted securities. At the end of 1963 over 3,300 firms listed with the SEC dealt in unlisted securities. The over-the-counter market differs from the organized exchange in many ways. There is no fixed schedule of fees for buying and selling securities. Dealers more frequently buy and sell from inventories and derive their profit from markups on this inventory.

Dealers and brokers are informally organized in the National Association of Security Dealers (NASD). Under pressure from the SEC, the NASD began to publish in February 1965 wholesale or “inter-dealer” quotations for the 1,300 issues on its national list. Accurate information on retail prices is not yet available for the over-the-counter market.

The SEC estimates that as of December 31, 1962, about 4,500 stocks in about 4,100 domestic companies were quoted only in the over-the-counter market (U.S. Securities and Exchange Commission 1963a). These stocks had an aggregate value of about \$90,000 million, almost half being constituted by stocks in banks and insurance companies. The value of stocks traded over-the-counter was only about one-fourth of those traded on the New York Stock Exchange but far exceeded the value of stocks traded on any other American exchange.

The volume of stock trading over-the-counter is estimated to have grown about eightfold between 1949 and 1961, reaching \$40,000 million in the



latter year. Although most of the very large companies are traded on listed exchanges, this is not true of stocks in banks, insurance companies, and mutual funds, which are frequently large and seldom listed. (In 1964, a Disclosure Act was passed by Congress which may substantially increase the listing of stocks in banks and insurance companies.)

In recent years a so-called third market has increased in importance. This market involves trading in shares listed on exchanges, but the trading takes place in the over-the-counter market rather than through the exchanges. The only definitive estimate of the volume of such trading indicates that in 1961 about \$2,000 million worth of stocks listed on the New York Stock Exchange were traded over-the-counter. This constituted almost 4 per cent of trading on the New York Stock Exchange. This percentage is undoubtedly increasing. The third market is growing in importance because the brokerage costs on the exchanges are fixed and do not permit quantity discounts, despite the lower cost per share of handling large orders. The over-the-counter market has no fixed schedule of fees and thereby permits large buyers and sellers to consummate transactions at lower cost than is possible on the exchanges. The increased importance of institutional investors, who frequently trade in large volumes, has contributed to the growth of the third market.

Outside the United States, most trading in securities is probably through exchanges, but some

**Table 1 — Distribution of total market value of individually held stock, by income class, 1960**

Adjusted gross income <sup>a</sup>	Per cent of total market value <sup>b</sup>
Less than \$5,000	12.4
\$5,000 — \$9,999	13.9
\$10,000 — \$14,999	11.4
\$15,000 — \$24,999	14.0
\$25,000 — \$49,999	15.9
\$50,000 — \$99,999	12.9
\$100,000 — \$149,999	4.8
\$150,000 — \$199,999	3.3
\$200,000 and over	11.5
<b>Total</b>	<b>100.1</b>

<sup>a</sup> Excludes income not subject to federal income tax.

<sup>b</sup> Percentages do not add to 100 because of rounding.

Source: Crockett & Friend 1963, p. 156.

is over-the-counter and through institutions such as commercial banks. In most foreign countries, listing requirements and permission to trade on exchanges are less restrictive than in the United States.

**Share ownership.** In the United States the number of individuals owning stock directly has risen from about 6 million in 1954 to about 20 million in 1966. This ownership is widely distributed among all income groups, as is indicated by Table 1.

Somewhat more is known about the importance of stock ownership by institutions. Such ownership has increased rapidly both absolutely and as a percentage of all stocks, as is indicated by Table 2.

Information regarding share ownership in 60 different countries is contained in a study by the

**Table 2 — Estimated holdings of NYSE-listed stocks by financial institutions and relation to total value (billions of dollars)**

Type of institution	1949	1956	1961	1962	1963	1964	1965
<b>Insurance companies:</b>							
Life	1.1	2.3	4.0	4.1	4.6	5.3	6.4
Non-life	1.7	4.5	7.7	7.1	8.2	9.5	10.1
<b>Investment companies:</b>							
Open-end	1.4	7.1	17.2	15.4	18.6	21.8	26.5
Closed-end	1.6	4.0	5.6	5.3	5.7	6.6	5.6
<b>Non-insured pension funds:</b>							
Corporate	0.5	5.3	18.7	17.9	22.6	27.5	32.5
Other private	•	0.4	1.1	1.0	1.3	1.6	1.9
State and local government	•	0.2	0.7	0.8	1.1	1.5	1.9
<b>Nonprofit institutions:</b>							
College and university endowments	1.1	2.4	3.7	3.3	4.0	4.5	5.1
Foundations	1.1	4.1	7.2	6.7	8.3	9.5	10.1
Other	1.0	3.1	5.6	5.0	5.9	6.8	7.7
<b>Common trust funds</b>	•	1.0	1.9	1.7	2.2	2.6	3.2
<b>Mutual savings banks</b>	0.2	0.2	0.3	0.4	0.4	0.4	0.5
<b>Total</b>	<b>9.7</b>	<b>34.6</b>	<b>73.7</b>	<b>68.7</b>	<b>82.9</b>	<b>97.6</b>	<b>111.5</b>
<b>Total market value of all NYSE-listed stock</b>	<b>76.3</b>	<b>219.2</b>	<b>387.8</b>	<b>345.8</b>	<b>411.3</b>	<b>474.3</b>	<b>537.5</b>
<b>Estimated per cent held by all financial institutions</b>	<b>12.7%</b>	<b>15.8%</b>	<b>19.0%</b>	<b>19.9%</b>	<b>20.2%</b>	<b>20.6%</b>	<b>20.7%</b>

• Less than \$50 million.

Source: New York Stock Exchange 1967.

New York Stock Exchange (1962). Three facts stand out: (1) there is no information about share ownership in most countries; (2) for all countries for which information is available, the number of shareholders is growing rapidly; and (3) share owners constitute a larger portion of the populations of western European countries, Japan, and the United States than of other sections of the world.

In Great Britain, it is estimated by the Wider Share Ownership Council that at the end of 1963 there were about 3.5 million individuals owning stocks directly and that about two-thirds of all adults of the United Kingdom owned stocks indirectly through various institutions. Institutional investors are of greater relative importance in Great Britain than in the United States. British institutions own about 50 per cent of all the securities listed on The Stock Exchange (of London) while the comparable figure for American institutions on the New York Stock Exchange is about 20 per cent. The greater importance of institutions in the United Kingdom is accounted for primarily by the fact that their insurance companies are not so narrowly restricted with respect to the ownership of shares.

Japan had about 4.5 million individual share owners in 1963, over eight times as many as were estimated for 1946.

**The regulation of securities markets.** In the United States, securities markets are regulated to some extent by state laws, but by far the most important regulation stems from the following federal statutes:

- (1) The Securities Act of 1933.
- (2) The Securities Exchange Act of 1934.
- (3) The Public Utility Holding Company Act of 1935.
- (4) The Investment Company Act and the Investment Advisers Act, both of which became effective in 1940.
- (5) The Securities Act of 1964.

(For an extensive bibliography on the first four of these statutes, their amendments, and judicial decisions and administrative rulings derived from them, see Dunton 1963.)

Provisions of the federal statutes are administered by the SEC, which was established in 1934. The Securities Act of 1933 requires the disclosure of financial information judged by the SEC to be relevant for the evaluation by investors of the prospects of corporations issuing new securities. The Securities Exchange Act of 1934 regulates securities markets and the operations of security brokers and dealers. It requires registration of securities exchanges and of securities traded on such ex-

changes; it restricts borrowing by brokers and dealers; it prohibits manipulation of prices of securities; it requires periodic filing of reports on listed securities and the provision of specified information and statements soliciting proxies. Exchanges make rules governing the conduct of their members, and the SEC passes upon the adequacy of such regulation.

The over-the-counter market was brought within the purview of the SEC through a section of the Securities Exchange Act authorizing national associations of security dealers to regulate their members. The NASD, authorized under the Securities Exchange Act, has an extensive set of rules governing the conduct of its members. The Securities Act of 1964 extends to the larger over-the-counter issues the regulations governing registration of securities, financial reporting, and the issuance of proxies that apply to issues listed on exchanges. About 2,000 over-the-counter issues are affected.

The Public Utility Holding Company Act of 1935 was designed to eliminate or control public utility holding companies that either operated in or sold securities in interstate commerce.

The Investment Company Act and the Investment Advisers Act provide for the registration and regulation of investment companies and investment advisers in order to insure a code of conduct and a level of financial responsibility acceptable to the SEC.

Since enactment of the Securities Act of 1933 and the Securities Exchange Act of 1934, the dissemination of financial information by corporations has greatly increased, and the practices of firms buying and selling securities or advising investors have changed significantly. Much of this change has taken place through self-regulation by the organized exchanges and through the NASD.

The impact of these changes and the adequacy of existing regulation were subjected to exhaustive study by a special group established by the SEC, which the Congress had directed to conduct such a study. The study (U.S. Securities and Exchange Commission 1963b) consists of five volumes covering the operation of exchanges and the adequacy of existing regulation, the over-the-counter market, the qualifications of persons in the security industry, security credit, mutual funds, and other matters. A great deal of information, much of it quantitative, is contained in the so-called *Special Study*, but there is controversy as to whether there was an adequate factual basis for the recommendations that were made (Stigler 1964a; 1964b; Friend & Herman 1964; Robbins & Werner 1964).



One major change resulting from the *Special Study* was to subject the over-the-counter market to regulation more like that imposed upon formal exchanges. For the formal exchanges, the major changes, in addition to those resulting from the Securities Act of 1964, were modifications in the regulations concerning floor traders and specialists. The intent of these new regulations was to reduce the advantage of floor traders and specialists relative to the trading public and to decrease possible conflict of interest between their role as speculators and their role of providing liquidity. The new regulations reduce the opportunity of specialists to profit from their access to "book information" about bids and offers, and to execute orders on their own behalf prior to orders on behalf of the public.

**Security credit.** Stocks and bonds are widely used as collateral for loans. At the end of 1964, such loans exceeded \$5,000 million in the United States. This was less than one per cent of the value of all securities listed on the New York Stock Exchange. Loans based upon collateral consisting of securities are subject to special regulation in the United States. The Securities Exchange Act of 1934 required the Board of Governors of the Federal Reserve System to issue rules and regulations concerning security credit, and Regulation T accordingly was issued by them and became effective October 1, 1934. Regulation T specifies the amount of credit that can be extended by security brokers and dealers relative to the value of the collateral. At the end of 1964, such loans were limited to 30 per cent of the value of securities at the time that the loan was made. On May 1, 1936, the Board of Governors of the Federal Reserve System issued Regulation U relating to loans by banks based on collateral in the form of securities. (Extensive information about security credit can be found in Bogen & Kroos 1960.)

**Security prices and business conditions.** Prices of stocks generally move in the same direction as business conditions, rising during periods of expansion and falling during recessions. Prices of common stocks are considered to be a "leading" indicator in that turning points in indexes of these prices typically precede turning points in general business conditions (Moore 1961). The prices of bonds, too, generally move in the same direction as business conditions (Kessel 1965). While stock prices tend to lead general business conditions, movements in bond prices are roughly coincident with movements in business conditions.

Changes in the general price level have different effects upon the prices of stocks according to

the capital structure of the underlying corporation (Kessel 1956). The general principle which explains these variations is that corporations with a large amount of net debt per dollar of equity benefit most from inflation and suffer most from deflation. Net debt is defined as a corporation's monetary assets (cash, government securities, accounts receivable, and notes receivable) minus its liabilities (accounts payable, notes payable, mortgages, preferred stock, and bonds).

The relationship between changes in general business conditions and the prices of stocks is complex because the changes in each series affect each other. Changes in business conditions obviously affect the prosperity of corporations and hence the value of stocks in them. Conversely, changes in prices of stock affect the wealth of individuals and institutions and their willingness and capacity to spend money and to make investments and hence affect general business. From October 1929 to June 1932, the prices of stocks listed on the New York Stock Exchange declined by over 80 per cent, and this reduction of many billions of dollars in the assets of owners of stocks undoubtedly contributed to the great depression of that time.

JAMES H. LORIE

#### BIBLIOGRAPHY

- BOGEN, JULES I.; and KROOS, HERMAN E. 1960 *Security Credit: Its Economic Role and Regulation*. Englewood Cliffs, N.J.: Prentice-Hall.
- COOTNER, PAUL H. (editor) 1964 *The Random Character of Stock Market Prices*. Cambridge, Mass.: M.I.T. Press.
- CROCKETT, JEAN; and FRIEND, IRWIN 1963 *Characteristics of Stock Ownership*. Pages 146-168 in American Statistical Association, Business and Economic Statistics Section, *Proceedings*. Washington: The Association.
- DUNTON, CHESTER 1963 *Legal Bibliography on Federal Securities Regulation*. Mimeographed, U.S. Securities and Exchange Commission.
- FISHER, LAWRENCE 1965 Outcomes for "Random" Investments in Common Stocks Listed on the New York Stock Exchange. *Journal of Business* 38:149-161.
- FISHER, LAWRENCE; and LORIE, JAMES H. 1964 Rates of Return on Investments in Common Stocks. *Journal of Business* 37:1-21.
- FRIEND, IRWIN, and HERMAN, EDWARD S. 1964 The S.E.C. Through a Glass Darkly. *Journal of Business* 37:382-405.
- FRIEND, IRWIN et al 1958 *The Over-the-Counter Securities Markets*. New York: McGraw-Hill.
- KESSEL, REUBEN A. 1956 Inflation-caused Wealth Redistribution: A Test of a Hypothesis. *American Economic Review* 46:128-141.
- KESSEL, REUBEN A. 1965 *Cyclical Behavior of the Term Structure of Interest Rates*. National Bureau of Economic Research, Occasional Paper No. 91. New York: The Bureau.

- MERRETT, A. J.; and SYKES, ALLEN 1963 Return on Equities and Fixed Interest Securities: 1919-1963. *District Bank Review* (Spring Gardens, Manchester, England) [1963], no. 148:17-34.
- MOORE, GEOFFREY H. (editor) 1961 *Business Cycle Indicators*. 2 vols. National Bureau of Economic Research Studies in Business Cycles, No. 10. Princeton Univ. Press. → See especially Volume 1, pages 674 and 676.
- NEW YORK STOCK EXCHANGE (1962) 1963 *Individual Share-ownership Around the World*. 2d ed. New York: The Exchange.
- NEW YORK STOCK EXCHANGE, DEPARTMENT OF PUBLIC RELATIONS AND MARKET DEVELOPMENT 1967 *Fact Book*. New York: The Exchange.
- ROBBINS, SIDNEY; and WERNER, WALTER 1964 Professor Stigler Revisited. *Journal of Business* 37:406-413.
- SPRAY, DAVID E. (editor) 1964 *The Principal Stock Exchanges of the World: Their Operation, Structure, and Development*. Washington: International Economic Publishers.
- STIGLER, GEORGE J. 1964a Public Regulation of the Securities Markets. *Journal of Business* 37:117-142.
- STIGLER, GEORGE J. 1964b Comment. *Journal of Business* 37:414-422.
- U.S. SECURITIES AND EXCHANGE COMMISSION 1963a *Twenty-ninth Annual Report*. Washington: Government Printing Office.
- U.S. SECURITIES AND EXCHANGE COMMISSION 1963b *Special Study of Securities Markets: Report*. 88th Congress, 1st Session, H. Doc. No. 95, Parts 1-5. Washington: Government Printing Office.

## SÉE, HENRI

Henri Sée (1864-1936), French economic historian, was born at Saint-Brice (Seine-et-Oise). He attended the Lycée Henri IV in Paris and then the Sorbonne. In 1887 he became *agrégé d'histoire* and taught for several years in the provinces. After successfully defending his doctoral thesis on the policy of Louis XI toward the cities, he was appointed professor at the University of Rennes in 1893. He remained there until a serious illness compelled him to retire prematurely in 1920. He continued to do historical research for the remainder of his life.

The work of Sée has a remarkable unity, which derives from his conception of the aims and the methods of history. His teachers, including Fustel de Coulanges, Ernest Lavisse, and Charles Seignobos, had trained him always to respect the rules of historical criticism. Sée believed that history must be rigorously objective, eschewing, on the one hand, grand systems and philosophies, and on the other hand, the domination of bias. But Sée also realized the futility of a kind of erudition that consists of data increasingly scattered among compartmentalized specialties, no longer useful for the development of synthetic knowledge.

Sée's *Histoire économique de la France* (1939-1942) is the work of a historian who cannot imagine studying economic facts apart from the milieu in which they occur and evolve, and the same spirit prevails in his *L'évolution commerciale et industrielle de la France sous l'ancien régime* (1925a) and his *Esquisse d'une histoire du régime agraire en Europe aux XVIII<sup>e</sup> et XIX<sup>e</sup> siècles* (1921). As a historian who was ever aware of the complexity of the facts, he found the Marxist doctrine of historical materialism unacceptable: from his viewpoint its great defect was its unilateral character. Although he was faithful to the socialist ideal all his life, it was from the effort of human thought rather than from "the blind movement of economic forces" that he expected humanity to progress, and his historical research was designed to show some of the conditions of rational progress.

Like his teacher Fustel de Coulanges, Sée was primarily interested in collective entities and institutions, even though he felt that Fustel had given too short shrift to the actions of individuals and the importance of "accidents." He believed that history should provide knowledge of the institutions, beliefs, and customs of a society, its way of thinking, its dominant concerns, and its guiding ideas. Sée felt that the most certain way of making his contribution to history as the "science of human societies" was to devote himself to both social and economic history and the history of ideas. Neither political or military history nor the biographies of rulers serve to reveal peoples to themselves. The most important dynamic factors in the evolution of societies are economic necessities, which predominantly determine the feelings and actions of governments and of social classes; however, the economic factor does not explain everything, and economic interpretations of history by no means have the character of laws.

Sée was a prolific author, his writings comprising no less than 28 books and 194 articles. As economics editor of the *Revue historique* he performed an important service to his profession.

A. CHABERT

[See also HISTORY, article on ECONOMIC HISTORY; MARXIST SOCIOLOGY; and the biography of FUSTEL DE COULANGES.]

### WORKS BY SÉE

- 1901 *Les classes rurales et le régime domanial en France au moyen âge*. Paris: Giard & Brière.
- 1921 *Esquisse d'une histoire du régime agraire en Europe aux XVIII<sup>e</sup> et XIX<sup>e</sup> siècles*. Paris: Giard & Brière.
- 1925a *L'évolution commerciale et industrielle de la France sous l'ancien régime*. Paris: Giard & Brière.



- (1925b) 1931 *Economic and Social Conditions in France During the Eighteenth Century*. New York: Croft. → First published in French.
- (1926) 1931 *Modern Capitalism: Its Origin and Evolution*. 2d ed. New York: Adelphi. → First published in French.
- 1927 *La vie économique de la France sous la monarchie censitaire*. Paris: Alcan.
- (1927) 1929 *The Economic Interpretation of History*. New York: Adelphi. → First published as *Matérialisme historique et interprétation économique de l'histoire*.
- (1928) 1933 *Science et philosophie de l'histoire*. 2d ed., rev. Paris: Alcan.
- 1929 *Évolution et révolution*. Paris: Flammarion.
- (1939-1942) 1948-1951 *Histoire économique de la France*. 2 vols. 2d ed. Paris: Colln. → Volume 1: *Le moyen-âge et l'ancien régime*. Volume 2: *Les temps modernes: 1789-1914*.

## SUPPLEMENTARY BIBLIOGRAPHY

- BECK, EARL R. 1951 Henri Sée and "The New History." *Florida State University Studies* No. 4:21-35.
- KNIGHT, MELVIL M. 1956 Henri Sée. Pages 107-117 in *Architects and Craftsmen in History: Festschrift für Abbott Payson Usher*. Tübingen (Germany): Mohr.
- PARKER, HAROLD T. 1942 Henri Sée. Pages 444-476 in Bernadotte E. Schmitt (editor), *Some Historians of Modern Europe: Essays in Historiography*. Univ. of Chicago Press.

## SEGREGATION

Segregation is an institutionalized form of social distance expressed in physical separation. It signifies a convergence of physical and social space and is to be distinguished from other social forms which also structure social distance in spatial terms, as, for example, the elaborate patterns of deference in African societies with sacred kingship or the different levels of seating (reflecting caste status) among Singhalese castes. The latter regulate social relationships between persons in situations of contact; segregation refers primarily to the separation of persons and the avoidance of contact.

Systems of segregation vary in the *criteria* which distinguish the segregated groups, whether biological, cultural, and/or status, and in the situations, or roles, which are segregated. Segregated roles may vary in extensiveness (extreme separation being represented by indirect rule in colonial society), in type (the intimate roles of primary relationships or the more impersonal roles in secondary relationships), and in the relative level of the actors (for example, where equal-status contacts are proscribed but dominant-subordinate contacts are fostered, as in racial segregation, or apartheid, in South Africa). Segregative systems may also be distinguished as compulsory or voluntary, as deliberate or spontaneous, and as influenced posi-

tively by attraction or negatively by disdain. Compulsory segregation involves deliberation and invidious distinction and serves the interests of those who impose it.

Although segregation imposes separation of persons and groups, it is by no means the antithesis of societal integration. The segregation of units may be a basis for integration, as in traditional Indian caste society, where a consensual basis for segregation was derived partly from shared religious values. Indeed, segregation may be conceived of as a general, although not a universal, aspect of social organization. It defines the boundaries between groups, locates the groups in the hierarchy, and regulates their interaction. Because of the intimate relationship of segregation to systems of domination, it is often highly resistant to change and readily becomes a focus of political conflict.

**Sociological approaches.** Early sociological concern with segregation derived in part from interest in the processes of social interaction (segregation being conceived of as "dissociative," in terms of a model of "attraction-repulsion" or "association-dissociation") but mainly from ecological studies, under the influence of animal and plant ecologists, geographers, and the theory of social Darwinism. The initial emphasis on the biotic approach stressed the subsocial elements in human society, with the corollary, explicit or implied, that segregation was a natural phenomenon; there was the same implication in the concept of natural areas [see REGION].

By contrast, the cultural approach, which developed in criticism of the biotic, stresses the influence of *values* on the spatial distribution of urban populations, especially the role of prejudice and discrimination in the segregation of ethnic and racial groups. Sociologists share with social psychologists and political scientists an interest in the psychological consequences of segregation as it affects the processes of prejudice, consensus, and political cleavage. And this interest has been further stimulated by the struggle for desegregation in the United States—a major focus of the movement to secure effective civil rights for Negroes—and by the threat of racial civil war in South Africa, which has a regime of systematic racial and ethnic segregation.

**Social structural factors**

Segregation, being an aspect of social structure, varies in its incidence and nature with the type of society. It hardly exists in relatively undifferentiated societies, such as those of the Bushmen, and the term is not really applicable to the divisions

between kin groups in segmented societies, which constitute separate rather than segregated units, loosely federated in certain situations.

**Social differentiation.** While segregation is associated with social differentiation, there does not appear to be a precise relationship between the two. Segregation may be highly elaborated both in societies where the division of labor is not greatly advanced and in modern industrialized societies. Thus, segregation is a basic principle of organization in the village communities of Indian caste society, where it regulates residence, access to amenities, and contact between the castes and where it receives ritual reinforcement in patterns of avoidance sanctioned by the threat of pollution. It is found in many "preindustrial" cities in Europe, Asia, and the Middle East and in the Muslim cities of north Africa, where ethnic and religious particularism may be manifested in separate quarters of the city as well as in occupational specialization. And it is characteristic of the most industrialized metropolitan cities of North America, in which there is extensive segregation of Negroes in residence, worship, education, and social intercourse. Nor does it seem that progressive industrialization will necessarily dissolve ethnic and religious particularisms that are irrelevant to the industrial process and replace them by distinctions resting on more universal criteria of qualification and achievement. This may be the trend within industry, but there is a continuing relevance of ethnic and religious identification as bases of association in the cities and in the extensive movement of urban populations to the suburbs.

**Social stratification.** The relationship between segregation and rigidity of social structure is equally complex. Segregation is characteristic of societies with complex class systems, although the ecological patterns of class segregation vary and a central city location may distinguish the upper classes in one society, the lower classes in another. There may be extensive segregation of roles in a class system, where residential segregation lays the foundations for segregation in education, religion, and recreation. But the barriers are much more permeable than in caste systems; in fact, in these two systems there is a direct relationship between the rigidity of the system of stratification and the rigidity of the system of segregation. This relationship, however, does not hold for the slave plantations, in which the extreme of social distance was nevertheless compatible with intimate contact. Segregation is one of many forms of institutionalized social distance and not in itself a precise indicator of social structure.

**Pluralism.** Plural societies have an affinity for segregation, especially when the social cleavages between racial, ethnic, or tribal groups are associated with cultural differences, as in the colonial societies of Africa. The policies of the metropolitan powers imply different policies toward segregation. In theory, Africans who acquire Portuguese or French culture may move freely in the circles of the conquerors, acculturation bridging the social cleavages between them. The British theory of indirect rule, by contrast, maintains the separation between the races by a system of parallel institutions, which extends to the courts of law; legal and other dualisms help maintain cultural diversity, and acculturation does not qualify individual Africans to cross the barriers of segregation and racial cleavage. In South Africa, a system of parallel institutions and consequent segregation serve to foster nationalism among the Afrikaans-speaking white population and to segregate racial, ethnic, and tribal groups under the policy of apartheid.

#### Patterns of racial and ethnic segregation

Different historical circumstances influence the patterns of segregation between the races. For example, after the conquest of Algiers the French settled inside the depopulated city, whereas in the protectorate of Morocco, which was constituted by treaty with the sultan, they established separate cities adjacent to the Muslim cities (Le Tourneau 1957, pp. 112-113). Segregation is also affected by the extent and nature of administrative regulation. In British and French west Africa, racial segregation was social and customary; in the Belgian Congo and southern Africa, it was legally imposed and sanctioned. In South Africa it is systematized under town plans, which tend to reserve the core of the towns for whites and to move nonwhites to the periphery or beyond or to establish them in satellite towns.

**Colonial Africa.** Whatever the colonial policy, the historical conditions, or the administrative framework, the general pattern in colonial Africa is one of racial segregation, and in east Africa and South Africa it extends also to intermediate groups such as Indians. The generality of racial segregation in colonial Africa flows from the convergence of many different types of social distance, based on race, power, class, and culture. Inevitably, the lower classes on the fringes of urban life (the so-called *sous-proletariat*) who live in the improvised shanty towns, or *bidonvilles*, are recruited from the population of the colonized, and the residents of upper-class neighborhoods are the colonizers. Cul-



tural differences extend racial separation to many institutional spheres. Even in shared institutions (as in the conversion to Christianity), ideologies of domination or sentiments of superiority fuse with cultural and racial differences to create varied and intricate patterns of integration and segregation, including, in extreme cases, reinterpretation of the communion with Christ.

*Tribal segregation.* Tribal segregation in colonial society is less general and more varied in its forms. In the old cities of Africa, uncontrolled residential patterns permitted tribal concentrations. In the new cities, housing policy, control over the influx of migrants, and the demand for housing all affected the possibility of tribal segregation. In some areas it was discouraged by indifference to tribal affiliation in the allocation of housing; in others, though less often, it was the principle which governed the allocation. Sometimes housing policy and social conditions were sufficiently flexible to permit a measure of voluntary tribal segregation. The relations between tribal hinterland and city, tribal composition and relative concentration or dominance of different tribes, extent and organization of industry, and political structure all have a bearing on the incidence of tribal segregation. Cultural differences encourage the spread of tribal segregation into a variety of institutions, as shown for example in the establishment of separate mosques by many tribal groups in Freetown, Sierra Leone (Banton 1957, p. 137). In general, the acquisition by different tribesmen of elements of a common culture in the urban setting and through the schools, as well as their common subordination to the colonial power and the processes of restratification, reduce the relevance of tribal affiliation and provide new bases of association.

African independence may have different consequences for racial and tribal segregation. In the case of racial segregation, the policy of Africanization is designed to effect desegregation, at any rate in the more public spheres of life, indeed, it may carry the process to the point of an African imposition of exclusiveness. In the case of tribal segregation, there is the bond of common race to bring tribesmen together, and the cultural differences between the tribes are more readily bridged than those between the races. But under some conditions of social structure, that is, where organization of the tribal unit offers access to power, the political struggle after independence may give enhanced significance to tribal affiliation and association.

*United States.* Segregation in the United States has rested on such varied aspects of pluralism as

race, ethnic background, religion, and culture, or combinations of these, in association with class differences. It is most resistant to change in the case of Negroes. Cultural differences between Negroes and native-born white Americans are not marked. The basis of segregation is the general interaction of lower-class status and invidiously defined racial status. Differences in racial status are institutionalized in the structure and culture of the society, although ambiguously, since this conflicts with the democratic ethos. Persons of low racial status have little access to opportunity for achievement, because of segregation in poor neighborhoods and inferior schools and concentration in manual and unskilled occupations. Conversely, the relatively lower levels of achievement reinforce the depreciation of the Negro's racial status.

The mechanisms of this racial segregation vary. In the South, the enactment of laws to enforce segregation was encouraged by the decision of the U.S. Supreme Court in *Plessy v. Ferguson* (163 U.S. 537, 1896), which held that the provision of separate but equal facilities was not a violation of the right to equal protection of the law under the fourteenth amendment to the constitution. But compulsory segregation did not include racial zoning (in contrast to South Africa), since the Supreme Court refused to accept the doctrine of "separate but equal" as legitimizing residential segregation. In the North, convention, restrictive covenants, discriminatory housing policies, poor wages, and the continued movement of new migrants into low rental areas of established Negro settlement all contributed to a high level of residential segregation, which provided the core for extensive institutional segregation in education, religion, and politics. With the removal of many of the legal supports of racial segregation, pursuant to decisions of the Supreme Court, and despite the official policies of the federal government and a number of states and cities, segregation is still carried forward by various deliberate means, such as the controls applied by boards of realtors, as well as by its own momentum and the consequences of past segregation. Since Negroes were largely excluded from the movement to the suburbs, the central core cities in the great metropolitan areas become increasingly nonwhite as the racially segregated urban populations receive the new Negro and other colored migrants.

*Ethnic groups.* The patterns of ethnic segregation in the United States are similar to those of racial segregation. Foreign migrants establish themselves in central city areas, in zones of transition, and in the interstices of the city, where they

create "Little Sicilies" or Jewish ghettos or other incapsulated communities, which are often characterized by poverty and social disorganization. In this respect they are very similar to the racially segregated communities; however, the ethnic groups and their settlements dissolve more readily in the environing society under the conditions of immense growth and vastly expanding opportunities experienced in the United States. Cultural differences diminish as the political and economic institutions, the schools, and mass media powerfully shape a common culture. With the passing of time, ethnic background tends to become detached from social class and culture. It still remains relevant for association or exclusion, and ethnic enclaves persist and are transported to new suburbs. But ethnic segregation becomes less general than racial segregation, less discriminatory, and less extensive in the range of roles affected (being emphasized more in primary than secondary relationships). Some ethnic identities have merged into an upper stratum of Anglo-Saxon and north European American identity. Religious identity—that of Protestant, Roman Catholic, and Jew—subsumes in some respects the ethnic identity, replacing its exclusive particularism with a more general principle of association.

**Comparison of racial and cultural criteria.** Comparison of Negro and Puerto Rican segregation in New York City, as it develops over the years, may provide some measure of the relative emphasis on racial and ethnic background in the integration of groups into North American society; however, interpretation is rendered difficult by an admixture of color in the Puerto Rican migration, and by the very different significance, history, and organization of the two groups in the wider society. Probably the situation of Negroes in the United States is unique, and their former slave status may be as relevant as their racial origin.

In England, there are appreciable numbers of colored people who entered the country after World War II; although they constitute little more than 1 per cent of the population, their presence has created an awareness of a color problem. Since they are of varied cultural background (mainly from India, Pakistan, and the West Indies), comparative analysis may yield some insight into both the role of race and of cultural differences in the exclusion and segregation of minority groups in England. If English prejudice against people of other races is only an expression of a more general xenophobia, and not specifically racial prejudice, as some English sociologists maintain, then it follows that cultural difference should be the crucial

variable in the exclusion of strangers. The evidence indicates some residential concentration of immigrant groups, whether colored or not, in different boroughs of London, and some dense clusters of colored groups, but also much dispersion. There is exclusion of colored people in primary relationships, much discrimination in rented accommodations, and some discrimination in employment (Glass 1960; Davison 1963; Deakin 1964). But citizenship rights may be fully exercised, regardless of race, and such public amenities as transport are not segregated.

The relations between race, culture, and segregation are, no doubt, indeterminate; they vary in different social contexts and historical situations. Segregation is by no means a universal corollary of racial differentiation. Racial difference may be relevant for social stratification, but expressed little, or not at all, in segregation; or its relevance may be indirect, that is, resting on association with class and cultural differences (the seemingly biological description of groups referring to sociocultural distinctions, as in Mexico); or the reality of racial difference may be submerged in a social myth of racial homogeneity, as in Argentina (Beals 1955; Taeuber & Taeuber 1965; Schnore & Evenson 1966).

Since racial segregation rests on relatively enduring characteristics, it is likely to be highly persistent, once it becomes established and linked with structures of domination. But even this is not certain. Residential and other forms of segregation by race may rapidly disappear with sharp changes in the structure of power, as is to be anticipated in the independent states of Africa.

### Ideological bases of segregation

After World War II, movements for racial desegregation were stimulated both by changes in the internal structure of the societies in which they arose and by the changed relationship between white and nonwhite peoples in the world structure of power. The emancipation of colonial peoples is essentially an emancipation of nonwhites from white domination. Concerned with eliminating the racial discrimination they suffer in their own countries, they are affronted by its practice in other countries. Since the distribution of races crosses national boundaries, movements for racial desegregation acquire international significance. It becomes increasingly difficult to justify racial segregation as *ideologies* of segregation are undermined by a world ethos of universal human rights.

**Religious doctrines.** Ideologies of segregation form around major social values. They may be



expressed in religious terms. Thus the concept of *pollution* regulates contact and avoidance in a most extensive range of roles by supernatural sanctions and by sentiments of deep revulsion for defiling contacts. Doctrines of *predestination* afford justification for controlling level of contact and for discriminating against persons who are not members of the elect; they may also serve for an extensive regulation of roles.

Resentment of invidious status under theories of predestination may provoke countertheories which emphasize reversal of status. In some of the African Zionist and messianic movements, it is the black prophet who seeks a "New Jerusalem" on earth for the chosen African people or closes the gates of heaven against the white man. In the racially exclusive religious movement known as the Nation of Islam, American Negroes may find eschatological vindication of their race and a rationale in divine destiny for a policy of territorial separation (Lincoln 1961).

Doctrines of racial predestination are not easily maintained in Christian churches, where they are almost universally disavowed. They may reappear, however, in doctrines which enjoin separation, although they abandon the invidious racial distinctions in deference to the world ethos of human equality. In one variant, the existence of separate races is taken as evidence of "divine providence" and of a religious duty to preserve racial separation by avoidance of contacts which might lead to the intermingling of races and loss of racial purity.

**Natural instinct theory of segregation.** Scientific ideologies rest on assumptions as to the nature of man and of human society. The assumed ubiquity of racial segregation is often taken as evidence that it derives from the instinctive nature of man. There are similar implications, though without racial connotation, in the concept of the fight for "turf" as the human counterpart of the animal instinct for "territory." A natural instinct theory of racial segregation may serve as an ideology for compulsory segregation, which is thus conceived as consonant with the true nature of man; or it may serve as a defense against compulsory segregation, since the natural proclivity of races to segregate themselves would render compulsion redundant.

**Measurement of segregation.** The prevalence of racial segregation cannot be doubted, although there is little precise information as to its extent. Segregation may be measured statistically by an index of dissimilarity in the spatial distribution of two groups, or by indexes of segregation, which calculate either the difference between the distri-

bution of one group and all other groups or the deviation from a postulated norm of even distribution. Moreover, an index of segregation may be combined with other measures, such as social rank and urbanization, to yield an urban typology, as in "social area analysis" (Shevsky & Bell 1955).

There are difficulties in the use of these measures of segregation. They are relative to the system of area units used, whether blocks or census tracts; they leave out of account the presence of segregated clusters within units, since they are based on the assumed homogeneity of these units; they may be affected by the relative proportion of the segregated group in the total population; and they measure the results of a complex of factors, not only the influence of the racial factor in segregation (Duncan & Duncan 1955, pp. 215-217). Furthermore, they impose the need for surveys to collect the basic data in countries which do not take a census of their populations or do not inquire into racial identity. But they do provide a basis for comparing racial segregation in different societies and for estimating its extent. In the absence of reliable comparative data from countries other than the United States, in which these measures have been widely used, assertions as to the generality and extent of racial segregation rest on impression.

**Other forms of ideological defense.** Other forms of scientific ideology postulate the inevitability of conflict in the contact between peoples of different culture or race and hence the necessity for avoiding contact. Or they assert as a corollary of the natural instinct theory of segregation, that racially mixed areas are inherently unstable and tend to revert to the segregated state. In this conception, when members of a subordinate race "invade" the residential area of the dominant race, the residential cycle ends in a racially segregated area for the invading group, since the original residents evacuate their homes; thus invasion becomes succession. In capitalist society, the protection of property rights and values may provide an ideological defense against desegregation. Prohibition of racial segregation in the leasing or selling of property may be represented as an infringement or diminution of rights of ownership and desegregation as a threat to property values under an assumption of inevitable depreciation.

Refutation of the theoretical assumptions in these ideologies involves the demonstration of contrary cases of stable racial intermingling, of contact without conflict, and of the maintenance, and indeed enhancement, of property values. This type of analysis rejects categorical assumptions as to

the nature of man and society and focuses on the conditions which give rise to racial segregation or those which favor the intermingling of races or promote desegregation.

# Segregation and prejudice

Desegregation may be effected by revolutionary changes in the structure of the society, as in many parts of Africa, or it may be sought within the existing structure by action against prejudice and discrimination. Segregation is not necessarily associated with prejudice, since it may be maintained by the institutions of society, independently of the attitudes of individuals. But in general, systems of segregation draw support from the prejudices of individuals and indeed foster those prejudices. In situations of systematic racial segregation which permit contact only on a basis of inequality, the routine experience of racial superiority may be expected to intensify sentiments of superiority and sentiments supportive of segregation; these sentiments in turn seek expression in further segregation, so that the levels of both prejudice and segregation are raised by their mutual interaction. Conversely, under certain conditions the experience of contact on a basis of equality may undermine prejudice and weaken the support for segregation. Studies of interracial housing projects in the United States have shown the influence of residential proximity, official policy, and perceived social climate in creating more favorable attitudes (Deutsch & Collins 1951; Wilner et al. 1955). Action programs to promote racial harmony and equality may emphasize desegregation, since racial segregation is mostly overt and public and hence amenable to control by legislation and administrative regulation and since attitudes may be modified by changing the context and experience of race relations.

LEO KUPER

[Directly related are the entries HOUSING; MINORITIES; PREJUDICE; REGION. Other relevant material may be found in CASTE; ETHNIC GROUPS; SECTS AND CULTS; STATUS, SOCIAL; STRATIFICATION, SOCIAL, article on THE STRUCTURE OF STRATIFICATION SYSTEMS.]

## BIBLIOGRAPHY

- BANTON, MICHAEL 1957 *West African City: A Study of Tribal Life in Freetown*. Oxford Univ. Press.
- BEALS, RALPH L. 1955 Indian-Mestizo-White Relations in Spanish America. Pages 412-432 in Conference on Race Relations in World Perspective, Honolulu, 1954, *Race Relations in World Perspective: Papers* . . . Edited by Andrew W. Lind. Honolulu: Univ. of Hawaii Press.
- DAVISON, R. B. 1963 The Distribution of Immigrant Groups in London. *Race* 5, no. 2:56-59.
- DEAKIN, NICHOLAS 1964 Residential Segregation in Britain: A Comparative Note. *Race* 6, no. 1:18-26.
- DEUTSCH, MORTON; and COLLINS, MARY E. 1951 *Interracial Housing: A Psychological Evaluation of a Social Experiment*. Minneapolis: Univ. of Minnesota Press.
- DRAKE, ST. CLAIR; and CAYTON, HORACE R. (1945) 1962 *Black Metropolis: A Study of Negro Life in a Northern City*. 2 vols., rev. & enl. New York: Harcourt.
- DUNCAN, OTIS DUDLEY; CUZZORT, RAY P.; and DUNCAN, BEVERLY 1961 *Statistical Geography: Problems in Analyzing Areal Data*. New York: Free Press.
- DUNCAN, OTIS DUDLEY; and DUNCAN, BEVERLY 1955 A Methodological Analysis of Segregation Indexes. *American Sociological Review* 20:210-217.
- FREEMAN, LINTON C.; and PILGER, JOHN 1964 Segregation: A Micro-measure Based Upon Compactness. Syracuse University, Maxwell School of Citizenship and Public Affairs, *Residential Segregation Study* [1964], October 19.
- FREEMAN, LINTON C.; and PILGER, JOHN 1965 Segregation: A Micro-measure Based Upon Well-mixedness. Syracuse University, Maxwell School of Citizenship and Public Affairs, *Residential Segregation Study* [1965], February 12.
- GLASS, RUTH (1960) 1961 *London's Newcomers: The West Indian Migrants*. Centre for Urban Studies, University College, London, Report No. 1. Cambridge, Mass.: Harvard Univ. Press. → First published as *Newcomers: The West Indians in London*. Chapters 2 and 3 analyze geographical distribution and discrimination in housing.
- GLAZER, NATHAN; and MOYNIHAN, DANIEL P. 1963 *Beyond the Melting Pot: The Negroes, Puerto Ricans, Jews, Italians, and the Irish of New York City*. Cambridge, Mass.: M.I.T. Press.
- GORDON, MILTON M. 1961 *Assimilation in America: Theory and Reality*. *Dædalus* 90:263-285.
- GORDON, MILTON M. 1964 *Assimilation in American Life: The Role of Race, Religion, and National Origin*. New York: Oxford Univ. Press.
- GRIER, GEORGE; and GRIER, EUNICE 1964 Obstacles to Desegregation in America's Urban Areas. *Race* 6, no. 1:3-17.
- HODGKIN, THOMAS L. 1956 *Nationalism in Colonial Africa*. London: Muller. → Part II, Chapter 1 deals with the new towns of Africa.
- JOHNSON, CHARLES S. 1943 *Patterns of Negro Segregation*. New York: Harper.
- KUPER, LEO; WATTS, HILSTAN; and DAVIES, RONALD L. 1958 *Durban: A Study in Racial Ecology*. London: Cape; New York: Columbia Univ. Press. → A study of compulsory segregation in South Africa.
- LE TOURNEAU, ROGER 1957 *Les villes musulmanes de l'Afrique du Nord*. Algiers: Maison des Livres.
- LIEBERSON, STANLEY 1963 *Ethnic Patterns in American Cities*. New York: Free Press.
- LINCOLN, CHARLES ERIC 1961 *The Black Muslims in America*. Boston: Beacon. → A paperback edition was published in 1962.
- MCENTIRE, DAVIS 1960 *Residence and Race: Final and Comprehensive Report to the Commission on Race and Housing*. Berkeley and Los Angeles: Univ. of California Press.
- OSOFSKY, GILBERT 1966 *Harlem: The Making of a Ghetto*. New York: Harper.
- PARK, ROBERT E. (1916-1939) 1952 *Human Communities: The City and Human Ecology*. Collected Papers, Vol. 2. Glencoe, Ill.: Free Press.



- PARK, ROBERT E.; BURGESS, ERNEST W.; and MCKENZIE, RODERICK D. 1925 *The City*. Univ. of Chicago Press.
- POLL, SOLOMON 1962 *The Hasidic Community of Williamsburg*. New York: Free Press.
- SCHNORE, LEO F. 1965 *The Urban Scene: Human Ecology and Demography*. New York: Free Press.
- SCHNORE, LEO F.; and EVENSON, PHILIP C. 1968 Segregation in Southern Cities. *American Journal of Sociology* 72: 58-67.
- SHEVSKY, ESHREF; and BELL, WENDELL 1955 *Social Area Analysis: Theory, Illustrative Application, and Computational Procedures*. Stanford Univ. Press.
- SJOBERG, GIDEON 1960 *The Preindustrial City: Past and Present*. Glencoe, Ill.: Free Press.
- SOUTHALL, AIDAN (editor) 1961 *Social Change in Modern Africa*. Oxford Univ. Press.
- TAEUBER, KARL E.; and TAEUBER, ALMA F. 1965 *Negroes in Cities: Residential Segregation and Neighborhood Change*. Chicago: Aldine.
- THEODORSON, GEORGE A. (editor) 1961 *Studies in Human Ecology*. Evanston, Ill.: Row, Peterson. → The ecological context of segregation.
- VAN DEN BERGHE, PIERRE L. 1960 Distance Mechanisms of Stratification. *Sociology and Social Research* 44: 155-164.
- WILNER, DANIEL M.; WALKLEY, ROSABELLE P.; and COOK, STUART W. 1955 *Human Relations in Interracial Housing: A Study of the Contact Hypothesis*. Minneapolis: Univ. of Minnesota Press.
- WIRTH, LOUIS 1928 *The Ghetto*. Univ. of Chicago Press.

## SEIGNIORAL ECONOMY

See MANORIAL ECONOMY.

## SELECTION

See SCREENING AND SELECTION.

## SELF CONCEPT

Historically, the domain of investigation subsumed under "self concept" by psychologists and sociologists was a major concern of theologians, philosophers, political essayists, playwrights, and novelists. Since differences in defining the domain exist even today, it is hazardous to start with a definition without first preparing ground on which terms of the definition can be built. This task is undertaken here through a glance at the status of the self concept in contemporary psychology and through a brief report of representative conceptions by various writers, noting convergences and differences. With an adequate definition, developmental and experimental research can be reviewed to derive leads for a conception of self that is scientifically based, rather than one that suggests "my word against your word."

### Current status and historical background

Especially since the 1940s, problems related to self concept have surged forth as indispensable

and legitimate topics for scientific study in psychology and sociology. The interest in self can be appreciated better if we consider the years of its relative disuse, and even disrepute, in the mainstream of psychology between 1900 and 1940. These years correspond roughly to the period of dominant concern with asserting the scientific nature of psychology by shying away from any concepts associated with the philosophical origins of psychology. The brave new world of scientism was represented by the models of Wundt and Titchener and by the behaviorism of Watson. The first asserted that the prime task of scientific psychology was the discovery of mental elements and then the laws of their compounding; the second, that the scientific task was a search for elemental reflexes and principles of their linkages. Each model, in its own terminology, advocated its approach as the only way to lay solid foundations for the ultimate explanation of more complicated forms of behavior.

Even then, the picture was not entirely monolithic. There were always those in search of unifying, integrating concepts to handle problems of the consistency of the person and the continuity of this consistency over time. The memorable chapter on self by William James (1890) maintained its impact on psychology and social science even during the period in question. James analyzed the self in terms of its constituent parts, as the sum total of what the individual considers himself and his to possess, including his body, his traits, characteristics, abilities, aspirations, family, work, possessions, friends, and other social affiliations. James neatly boiled down the problem of the maintenance of self-esteem into a formula (Self-esteem = Success/Pretensions) that served as the basis of many formulations and much research by psychologists and sociologists. Among his other insights, James anticipated the core of modern formulations on reference groups. He wrote of the individual's "image in the eyes of his own 'set,' which exalts or condemns him as he conforms or not to certain requirements that may not be made of one in another walk of life" [1890, vol. 1, pp. 294-295, see also the biography of JAMES].

Among other influential contributors was James Mark Baldwin, who gave an interactionist account of self development, epitomized in his formulation: "The ego and the alter are . . . born together" ([1895] 1906, p. 338). Some years later, the development of self through social interaction was elaborated by the sociologists Charles H. Cooley (1902) and George Herbert Mead [1913; see also INTERACTION and the biographies of BALDWIN; COOLEY; MEAD].

Of course, even in the first half of the twentieth century, there were the "personalistic" psychologists, such as Mary W. Calkins and Wilhelm Stern, who insisted that self-reference was characteristic of all psychological activity. Such formulations would find an ardent advocate later in Prescott Lecky's *Self-consistency* (1945). But the experimental mainstream was a psychology that had banished the self and other integrative concepts. [Cf. Allport 1943; see also PERSONALITY: CONTEMPORARY VIEWPOINTS, *article on A UNIQUE AND OPEN SYSTEM*; and the biography of STERN.] As a result, it never came to grips effectively with problems of human motivation (e.g., Koch 1956) nor with regularities in human behavior on the conceptual level of functioning (Schneirla 1946; 1951). Research results were fragmentary and attempts to put the fragments together proved inconclusive and even contradictory, showing little resemblance to the characteristic consistency of the person as he pursues the satisfaction of his needs for food, sleep, and sex and as he works and plays with his fellows.

As a rule, the person searching for food is not guided by hunger alone; he is also guided by what he considers to be proper food, the place where it is located, and the atmosphere where it is eaten. The tired traveler looking for a hotel is not merely concerned that the bed be like his own bed at home or perhaps a bit more comfortable. He is equally, if not more, concerned with the class of the hotel, whether it is on a par with his standing in life, and how his friends, employer, or sweetheart will react when he announces where he is staying. Ordinarily, the search for a sex partner is not guided merely by the urge for any sex object immediately available, who will serve to reduce tension and restore the homeostatic level. The episodes of sexual activity, beginning with the search, are also guided by proving one's worth and status as a man or woman both to oneself and to others.

Therefore, the individual's self is involved in more than striving to raise one's prestige, status, or self-esteem. Typically, self becomes involved in the operation of motivational urges, whether the individual is aware of it or not. The pervasive presence of self factors is expressed well by Shoben: "In any case, self-involved behavior seems close to impossible to explain on the basis of a tension-reduction model, and the postulation of self-involvement seems necessary to account for the pursuit of long range goals so typical of human motivation" (1962, p. 771). [See MOTIVATION, *article on HUMAN MOTIVATION*.]

The regulation of behavior by self is not restricted to motivational activity. Self concerns

enter as regulating factors into many psychological processes: judging, perceiving, learning, remembering, thinking, planning, and decision making. In performing a task, level of performance is not determined solely by the nature of the problem (difficulty or ease, for example). The goals one has erected for oneself in general, the place of the particular task in one's scheme of goals, and one's standing relative to others on the task all enter the picture and affect one's performance. During his development, the individual comes to stand in established reciprocities with others, as high, low, or equal; friendly or unfriendly; dominant or subordinate. When stabilized, these reciprocities form patterns producing regularity and consistency in the individual's dealings with others.

In series of encounters with others, the individual develops self-identity, in various respects, which is reflected in consistency in his dealings with others and with situations from day to day. In different situations, relevant components of the self are aroused, lending to his behavior characteristic modes of reacting to, and coping with, situations.

In brief, the growing interest in a self concept reflects the search for integrative concepts, particularly in motivation, where empirical work has tended to be fragmentary. Studying motives separately has fallen short in providing an adequate account of human motivation. The self enters into the operation of human motives as a regulative factor. So, too, self enters into other psychological processes. Involvement of the self in these processes is reflected in the consistency of the person and its continuity from day to day. In fact, self involvement in particular aspects of the kaleidoscopic stimulus world is the basis for the experience of continuity in personal identity.

### Usages of self concept

For reasons mentioned, problems of self or ego came irrevocably to the foreground as a legitimate area of investigation. Yet exactly what the concept covers, its relation to other integrative concepts (e.g., role, ego, personality) are still not crystallized. A convenient summary of different usages is presented in Hall and Lindzey (1957, pp. 467-499).

Not including the primarily psychiatric and clinical conceptions in our consideration, a major definitional divergence is represented between those who differentiate between the concept of self and the concept of ego and those who use the concepts of self and ego interchangeably. Ausubel (1952), Chein (1944), Hilgard (1949), Murphy (1947), and Symonds (1951) are among those



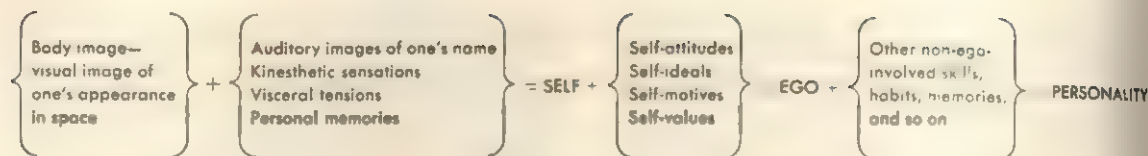


Figure 1

Source: Adapted from Ausubel, David P., *Ego Development and the Personality Disorders*, 1952, Grune & Stratton, Inc. Reproduced by permission.

who propose to use the concepts of self and ego in a differentiated way. Allport (1943), Snygg and Combs (1949), M. Sherif and Cantril (1947) are among those who use the words self and ego interchangeably.

**Self differentiated from ego.** An account of all the possible variations in differentiating the ego concept from the self concept is not within the scope of this article. It will suffice to examine representative ways in which they are differentiated. Murphy's definitions of self and ego in his monumental *Personality* (1947) provide one example: "Self: The individual as known to the individual" (p. 996); and "Ego: Group of activities concerned with enhancement and defense of the self" (p. 984). Here, "self" is used to mean the object consisting of so many attitudes and feelings in regard to the person himself, and "ego" is used to refer to the representative processes or activities.

Ausubel offers still another example. His diagram of the relationship of self, ego, and personality (Figure 1) clarifies his differentiation (1952, p. 13).

It may be fair to summarize Ausubel's schema by saying that the self is made up primarily of perceptual components, and ego consists of these and affectively charged conceptual components (self-ideals, self-values, etc.). But a clean separation between perceptual and conceptual components does not seem to fit into the ongoing developmental picture. As the child acquires labels and categories of language, "the self becomes less and less a pure perceptual object, and more and more a conceptual trait system" (Murphy 1947, p. 506).

Once the child starts acquiring language the body image becomes increasingly invested with value attributions that vary from culture to culture and from class to class (e.g., modesty in exposure, cleanliness, desirable body proportions, proper items to cover and adorn the body). Likewise, in building a self-picture the place of one's name is not restricted to auditory images: as McDougall noted, it soon becomes a handle to which many attributes are tied. Anthropologists have reported

cases in which personal names are changed at important transitions in people's lives. For example, A. R. Radcliffe-Brown (1922, p. 119) reported the change of name of Andamanese girls at the time of first menstruation. Likewise, Rivers reported the case of a Melanesian culture in which "on marriage both man and woman change their names and assume a common name" (1914, p. 347).

**Self interchangeable with ego.** Some authors seem to use either the concept of ego or the concept of self to cover a broad range. For example, Shoben defines "self" as "a relatively stable organization of values that mediates and focuses behavior" (1962, p. 771). George A. Miller defines "ego" as "the individual's conception of himself" (1962, p. 347). In view of this terminological malaise, it might be preferable to discard the labels "self" and "ego," fraught as they are with historical entanglements, and to use new terms not so encumbered. The new term proposed by Allport to replace the self and ego concepts is "proprium" (1961, p. 127). However, contemporary social psychologists, concerned as they are with the development of the personal identity of the individual—its enhancement and defense in social relations—seem to get along in their conceptualization without being concerned over the terminological differentiation between ego and self (Krech, Crutchfield, & Ballachey 1962; Newcomb, Turner, & Converse 1965; Secord & Backman 1964; Sherif & Sherif 1956).

### A characterization of self concept

On the basis of the foregoing considerations and surveys of theoretical and empirical literature, several propositions seem warranted.

Self (or ego) is not innate, as are the individual's organic urges—such as hunger and sex—which are subject to homeostatic regulation of the organism. Self (ego) is the product of interaction, from infancy onward, with the individual's physical and social environment. This interaction is associated with novel and familiar sensations: pain, resistance, acceptance, rejection, and gratification. The proposition that self is a developmental forma-

tion is one of the most widely documented generalizations among investigators from the nineteenth century on, despite variations in the specifics of their accounts.

There is also agreement that this psychological formation of the human individual is not a unitary structure, appearing full-blown. It develops first as a perceptual system based on the differentiation of one's body and its parts from its surroundings. It becomes progressively a more complex conceptual system, consisting of evaluative categories with associated traits or attributes. New classifications and qualities (e.g., adolescence, old age, acquisition of new interpersonal roles and social status) continue to be incorporated in the system during the life span (cf. Murphy 1947, pp. 489 ff.; Snygg & Combs 1949, pp. 78-79).

To emphasize the multifaceted character of the person's identity, William James (1890) analyzed this uniquely human formation in terms of the person's "selves." However, various components of the self formation are interrelated; different components may be mutually supportive or conflicting, depending on the situation. To underscore both the multifaceted nature and interrelated nature of self components, it is preferable to use a generic concept (self) subsuming the components, referring to the latter by terms other than the various "selves."

Therefore, the components are referred to as attitudes or self-attitudes, to specify that they pertain to relatedness of self to objects and persons important to the individual. Here, the term "attitude" is used to refer to more or less lasting evaluative categories of the person. (More transitory internal states may be referred to by other terms, such as "set," "expectation," or "bodily urge.") The definition of "attitude" here is similar to that given by Smith, Bruner, and White: "We define an attitude as a predisposition to experience a class of objects in certain ways, with characteristic affect; to be motivated by this class of objects in characteristic ways; and to act with respect to these objects in a characteristic fashion" (1956, p. 33). [See ATTITUDES.]

Combining these fundamental points, the characterization of self (ego) is as follows: Self is a developmental formation in the psychological makeup of the individual, consisting of interrelated attitudes that the individual has acquired in relation to his own body and its parts, to his capacities, and to objects, persons, family, groups, social values, goals, and institutions, which define and regulate his relatedness to them in concrete situations and activities.

The attitudes that compose the self system are, therefore, the individual's cherished commitments, stands on particular issues, acceptances, rejections, reciprocal expectations (roles) in interpersonal and group relations, identifications with persons or values, and personal goals for the future. When any of these—singly or in varying combination—enter as factors in ongoing psychological activity, owing to their relevance, behavior acquires characteristic directionality (positive or negative), becomes more consistent and selective, and is less bound by the ups and downs of the immediate situation. Ego involvement simply means involvement of such an attitude in ongoing psychological activity. Ego-involved activity is characterized by highly sensitized mobilization of the individual's psychological processes (discrimination, perception, learning, remembering, problem solving, decision making, and so on). Ego-involved activity, revealing aspects of an individual's personal identity, has been demonstrated empirically in many studies since the 1940s. (Note again the terminologically unhappy fact that these studies are almost always presented as studies of ego involvement, not self involvement.)

Defining, as they do, the individual's identity relative to objects, persons, groups, and situations around him, these self-attitudes determine the individual's experience as an active agent, principal actor, or executor when they participate in ongoing psychological activity. Therefore, it is not parsimonious to posit separate psychological concepts to account for experience of the person as executor and doer (cf. Ausubel 1952, pp. 41-46).

Thus, the unique formation that takes shape as the self or ego of the human person consists of a set of components that define his stabilized bearings in relation to the physical and social surroundings. They define his ties, his reciprocal expectations of his rights and responsibilities, and his motivational-emotional claims relative to objects as well as other persons, groups, and institutions. They define his stands and commitments relative to the social issues that are important to him. All of these components of the individual's self set him apart as a person in diverse ways and capacities with a sense of unique self-identity.

The stability of the person's self-identity is, therefore, dependent on the stability of the ties, roles, commitments, and orientations that compose it. They provide anchorages relative to which he gauges his personally felt status, security, and prestige, and his maintenance or gain in position in the social scheme of which he is a part. His experience of success or failure arises relative to



them. Hence, maintenance and continuity of the stability of these anchorages from day to day become matters of utmost importance in the dynamics of his motivational scheme.

Disruption of the stable anchorages that define the person's bearings relative to his surroundings arouses tensions, referred to as the state of uncertainty or anxiety, that are distressing or even unbearable. Experimental evidence shows that the loss of stable anchorages in the person's surroundings arouses feelings of uncertainty and insecurity, causing him to flounder about in efforts to restore his sense of personal stability. The experience of loss of his bearings makes an individual more susceptible to conforming to social influences to which he would ordinarily pay less heed (M. Sherif & Harvey 1952).

Similarly, disruption of stabilized personal ties, loss of acceptance by groups that he values, or lack of stable ties with others generate intense feelings of aloneness and personal rejection. Communications that advocate points of view divergent from the strong commitments incorporated in the self system arouse dismay, irritation, and tension. The committed person finds such communication more divergent from his stand, more unreasonable, and more obnoxious than does a person who is less committed. Thus, self-attitudes are resistant to communication designed to change them.

### Formation of self (ego)

Based on extensive surveys of research on child development (M. Sherif & Cantril 1947; Murphy 1947), this account notes a sequence of major events marking ego formation. Differences in age of occurrence and forms are to be expected in differing cultural and socioeconomic conditions.

Neither early observations of individual children nor modern studies of groups of infants provide evidence for innate self-identity. The infant's behavior is readily attributable to bodily states (hunger or sleep, for example) or the duration of immediate environmental stimulation. In fact, consistent differentiation and localization of the body parts require repeated experiences in localizing stimulation, meeting obstacles arousing unpleasant sensations, and exploring both body parts and environment. The earliest manifestations of self-awareness appear during this "perceptual stage" (Murphy 1947), in which the body is perceived apart from the surroundings. [See INFANCY, *article on* INFANT DEVELOPMENT; SENSORY AND MOTOR DEVELOPMENT.]

Cases of prolonged isolation in childhood suggest that ego formation beyond the stage of a per-

ceived body depends upon interaction with other persons who are members of a social and linguistic community (cf. Lindesmith & Strauss 1949). This accords with early theoretical accounts and modern formulations based on research conceiving of self as a product of interaction with others (e.g., Wallon 1934; Piaget 1932).

Facilitations and resistances, acceptances and punishments by persons who care for the child are landmarks for the developing self. Through such interactions, the child begins to acquire spoken language, at first slowly, then at a rapidly accelerating pace (during the second or third year of life). The acquisition of language is the single event most responsible for producing the consistency in behavior that requires postulation of a self system. [See LANGUAGE, *article on* LANGUAGE DEVELOPMENT.]

When the child is able to designate self and others through personal pronouns, a name, or verb forms, even the bodily self is delineated more sharply from its surroundings. Henceforth, the body itself is classified and endowed with attributes (favorable or unfavorable) through interaction on a verbal level. The striking variations in boundaries and properties of body image found in persons of different sex, age, culture, and states of bodily injury led Fisher and Cleveland (1958, p. 367) to postulate that the body image is best conceived as "a representation of attitudes and expectancy systems" related to the body and other persons. [See BODY IMAGE.]

Consistency in behavior is greatly enhanced through verbal interaction and conceptual classification, as research findings on the appearance of discriminatory responses by and toward American Negro children show so clearly (Clark & Clark 1947; Horowitz 1944). Very young children responded to skin-color differences but not preferentially. However, in a society systematically discriminating against Negroes, white and Negro children exhibited consistent preferences for light skin color by the age of 5.

While not every aspect of linguistic behavior is ego-involving, the child's instrumental mastery of language has general results that apply equally to the formation of categorical and attributive links between self and others (Lewis 1963). It liberates behavior from the total dictation of fluctuation in momentary bodily needs and the physical variations in external stimulation. From extensive research with children and cases of brain damage, Lurila concluded that "adoption of a verbal rule at once modifies the nature of all subsequent reactions . . . the stimulus in question becomes not a mere signal but an *item of general information*,

and all subsequent reactions depend more on the *system* it [stimulus] is taken into than on its physical properties" (1961, p. 44, italics in original).

The family and other adults in charge of the child's routine exert profound influence through their words and deeds on the nature and quality of the classificatory schemes defining what he is and is not. However, in line with G. H. Mead's theoretical scheme, the research of Piaget (1932) demonstrated that stability in the child's relatedness is attained slowly and, most decisively, in the company of other children. Little children behave consistently in dealings with others, according to rules or dictums handed down by authority figures, but easily lapse from them in the absence of authority. By participating in play with other children, the child has the greater opportunity to grasp notions of reciprocity with others and of mutual adherence to rules governing the give and take. [See DEVELOPMENTAL PSYCHOLOGY, article on A THEORY OF DEVELOPMENT.]

Thus, a series of experiments and "age-norm" studies (reviewed in Sherif & Sherif 1956) reveal that consistency in competing with others, in co-operating with others, in expressing sympathy at another's distress, in responsibility for self and others, and in setting goals for one's own performance appear gradually as the child participates in social and cooperative forms of play in contrast to earlier side-by-side or parallel play.

Slowly—and with confusion—the classificatory schemes defining self and others acquire longer dimensions in time and space. The immediate is located with reference to increasingly greater ranges of past and future (Lewis 1963). Geographical ranges of self-relatedness increase gradually from the perceptual scope of the child ("my house") to more distant and unseen, but conceptualized, places and the people in them.

The scope of time, place, and kind of persons included in categories for self and not-self varies enormously, along with their qualitative nature, according to prevailing social, cultural, material, and socioeconomic environments. The common property of self components in socialized members of all societies is the regulation of behavior within bounds of acceptability defined by the particular organizational and value systems of those groups with which the person identifies himself. Consistent evaluations of one class of objects (persons) and differential (sometimes invidious) evaluations of others is evidence of an ego component—an attitude relating self to that stimulus domain. The conception of self components (attitudes) as categorical schemes for evalua-

tion, including a latitude for acceptance and a latitude of rejection by the person, has led directly to operational procedures for assessing the person's ego involvement in various respects (Sherif, Sherif, & Nebergall 1965; Sherif, Sherif, & Kent 1967).

In addition, all known cultures present periods or stages in human development requiring alteration of the self concept as it is formed at the time. The years of transition from childhood to adulthood and the years of old age are two such periods. These particular transitions are marked by bodily changes that impel a change in self.

### Self and reference groups or sets

Research attempts failing to include explicit environmental factors have yielded inconclusive correlations between the person's self-ratings at two different points in time and subjectively circular accounts of self (Wylie 1961). In analyzing the "presentation of the self" in a variety of social situations, Goffman concluded that the person's characteristic consistency is a "collaborative manufacture" involving social arrangements and supporting performances by people within them (e.g., 1956, p. 253). Even advocates of a "phenomenal self" as determinant of all behavior could not ignore that much personal consistency and many changes in personal identity that occur with age or with taking on occupations and different social status are directly related to stabilities and changes in the social environment.

But individuals are differentially selective of their social environments, and modern social systems are extraordinarily complex and changing. A particular individual could develop within varying combinations of groups and institutions with differing values, rates of change, and stability. Owing to his advanced conceptual abilities, man may relate himself to persons and groups not present, spatially or temporally.

Therefore, unless the person is a fully integrated member of only one group and unless the social situation is fully structured as to socially desired outcomes, there is a need for concepts relating self to environmental events that may not be contemporaneous or spatially present. One such concept is that of reference groups or sets; these are defined as groups or classifications of people to whom the individual relates himself psychologically or to which he aspires to belong (Sherif & Sherif 1956; see the 1948 edition).

The concept of reference group, linking self components to environmental structures not necessarily present, is a distinctly human concept,



as the student of animal behavior J. P. Scott observed (1953, p. 69). It is unnecessary in studying subhuman behavior. It is probably unnecessary for analysis of self in an isolated, illiterate group. Only in differentiated societies is it possible that groups of actual membership may be anchors for self-identity.

From the research viewpoint, the concepts of self as a constellation of attitudes linked with identifiable reference groups and sets provide tools for integrating behavioral (individual) and sociocultural levels of analysis (Sherif 1962). A sharp dichotomy between individualistic and cultural (group) determination of behavior resists both argument and analysis but yields to research evidence that groups and populations have properties related to the development of self-other attitudes; self-other attitudes, in turn, display dimensional and classificatory features not necessarily identical with sociocultural prescriptions, but greatly clarified when related to them.

**Reference groups and self appraisal.** Asked to answer "Who am I?" the person's usual response is his name and social classifications, sometimes modified by adjectives placing him within the category (cf. Bugental & Zelen 1950; Kuhn & McPartland 1954). The social labels are typically exhausted before idiosyncratic responses appear.

The psychological import of this rather prosaic finding is vast: Each social category carries a complex of characteristics and positions within it that etch the specifics of a self image whether the individual actually belongs to that category or not. The bulk of the population of the United States, identifying themselves as "middle class," display a constellation of attitudes related to a middle-class position whether their individual socioeconomic levels warrant the designation or not (Centers 1947). Female college students who come from conservative families and, in time, take a liberal college student body as their reference group change their attitudes in diverse respects, not merely in identifying themselves with a particular college. However, those retaining a primary group reference, while classifying themselves as students, do not change at all (Newcomb 1950).

Groups, populations, and sets of people live in differing material and social environments that give focus to a complex of social values, setting bounds for the self-radius of persons who classify themselves as "belonging" within them. Thus, in the United States, the conceived range of achievement for self (from minimum to maximum) and the nature of personal goals varies systematically according to socioeconomic level and sociocultural

composition of reference sets; the range of achievement is most similar with respect to certain material possessions uniformly portrayed as success symbols in mass communication (Sherif & Sherif 1964). While upper-class youth are more likely to have high self-esteem than lower-class youth, the largest differences in self-esteem occur between groupings differentiated by both class and distinctive religious-ethnic cultures in the United States. Values cherished by youth in different socioeconomic and religious groupings differ, but there are marked similarities among them that indicate that they are all, nonetheless, members of the same society (Rosenberg 1965).

The reference-group concept permits specification of the particular group or set of people on whom the individual depends in appraising himself and others. In complex and changing societies, these people may not be readily identified by socioeconomic classification or geographical location. When the person locates himself within a set or group of people, the relative status of this group in the social organization and his own position within it serve as standards (anchors) for his appraisals of performance by himself and others.

In an experiment that served as a model for numerous others, Chapman and Volkmann (1939) showed that one's level of aspiration is anchored in the status of his own reference set relative to other reference sets. University students spontaneously raised or lowered their predictions for self when the group with whom they compared self was respectively "inferior" or "superior" to their own reference sets.

When anchored in the person's reference set, self-attitudes exert a stabilizing effect on performance. The relevant experiments have all used special instructions or varied tasks to assess the effects of ego involvement, and the findings hinge on the researcher's prior knowledge of values that are prized by the subjects' reference set (for example, for college students high intelligence, good university records, and contribution to science). The outcome of such experiments has been the repeated demonstration that such manipulations, arousing the person's mettle to prove his worth, result in greater consistency (less variation) of behavior in setting goals for performance, in self-confidence, in various personal characteristics, in learning, and in forgetting (see Allport 1943; M. Sherif & Cantril 1947). [See REFERENCE GROUPS.]

**Discrepant group membership.** While existence in an interacting group affects the individual's self concerns whether he takes stock of himself on their terms or not, attitudes relating self to others

are most predictable when the group is also the person's reference group. Siegel and Siegel (1957) found that over a period of time a person's attitudes coincided most with the values of groups that were also reference groups for him. A person who lived with a group whose values were at variance with his reference set was affected by it, but his attitudes resembled those of the reference set more closely.

Judgments of one's standing (rank) relative to others and of the rank assigned to self by others differ for those who are tied to the immediate group and those whose reference sets are outside of it. A study of small task groups in an isolated military base showed a general tendency to overestimate one's own rank and the rank that others would assign one. As a result, self-ranking and estimated ranking by others were more closely related than either were to actual ranks made by fellow members. However, those whose self-rankings disagreed most with the rank assigned them by others in the group were preponderantly married men with higher military rank and educational level. As a result, they were less tied to the reactions of the work group in rating themselves and tended to use standards from their other reference sets (Reeder, Donohue, & Biblarz 1962).

**Salience of different groups.** "I, who for the time have staked my all on being a psychologist, am mortified if others know much more psychology than I. But I am contented to wallow in the grossest ignorance of Greek. . . . Had I 'pretensions' to be a linguist, it would have been just the reverse" (James 1890, vol. 1, p. 310). The problem of self-reference-group relationship stated succinctly by James still holds the greatest promise for future research.

All components of the self are not equal in importance to the individual over time, and their relative importance varies in different situations. For every person, there is some order or hierarchy in the constellation of self values such that the ways, the time, and the relative frequency with which he becomes ego-involved differ systematically from person to person. Differences of this sort between individuals acquire some rational pattern when the hierarchy of their self-values is related to the relative importance of their various reference sets.

**Adolescence and marginality.** The period of adolescence in modern societies may serve as a prototype for study of the problem. During the prolonged and ill-defined transition from childhood to adulthood in Western industrialized societies, the adolescent is betwixt and between ref-

erence groups, much like a "marginal man." Lack of stable anchors for self arouses experiences of uncertainty and conflict. Grownups provide few ready-made paths for the transition that fully satisfy the adolescent's growing urges and his desire to prove himself adult. At this same time, he is provoked by the uncertainties of his standing. Scanning his environment for some stable anchor for self-identity, he turns increasingly toward age-mate groups and sets. Even though his parents be loved and valued, the result is reduced emphasis on parental capability and overestimation of the worth of age-mate capacities, as Prado (1958) has shown experimentally.

Absorption of self-identity within the organizational pattern and values developed in age-mate groups varies with the availability and stability of other reference ties (family, school, community, etc.); however, the adolescent's hierarchy of ego attitudes is closely linked to the values of his informal groups. When these values conflict with adult prescriptions, the defiance of adult standards, including legal violations, is self-justified by measuring oneself relative to the values of the youth group [Sherif & Sherif 1964; *see also* ADOLESCENCE].

**Multiple group membership.** Like many adults in multifaceted, complex societies, the adolescent belongs to multiple groups whose values may conflict. The member of multiple groups with contradictory values frequently faces situations of choice in abiding by the values of one reference set at the expense of the other. Still more painfully, he or she may form ego attitudes relative to different reference groups whose simultaneous arousal is bound to produce psychological conflict, uncertainty, and confused search for resolution. Such, for example, is the plight of many women today who conceive of themselves simultaneously as women and mothers in traditional terms and as modern, independent women on a par with men (Seward 1946).

Choice situations provide one index of the salience of the individual's self-attitudes. For example, during the crisis over desegregation in Little Rock, Arkansas, ministers faced the choice of speaking out against violence, as ministers of brotherly love, or of remaining silent in keeping with the standards of their segregationist congregations. Only a few of them chose to speak out in congruence with their ministerial self-identity (Campbell & Pettigrew 1959).

**Stability and change of ego attitudes.** Numerous instances of change in ego attitudes have been mentioned in this article. Stability may be at-



tributed in large part to the regularity and continuity of the person's social setting. Yet resistance to change inheres in their formation and functioning as well. Changing attitudes that form part of the person's relatedness to other persons, groups, and values amounts to changing part of the self-identity, which is the epicenter of experienced personal stability, even though it may not be an integrated harmonious structure. Disturbance of self-attitudes once they are formed is psychologically uncomfortable and even painful.

Resistance to change increases in proportion to the degree of personal commitment on the matter in question. When a person is highly involved in the matter in question, he is less likely to change his attitude readily and, when he does, is likely to change less than persons to whom the attitude is of moderate importance. In matters of high ego involvement, the individual is very discriminating in accepting proposals that differ even slightly from the values he has come to cherish as defining the attributes of himself and other persons who count in his eyes; he is prone to reject all those seen as different. Thus, his categorization of ego-involving material reveals broad latitudes of rejection in comparison with the narrow, but cherished, range of acceptance. Attempts to persuade him otherwise are likely to fail within the range that he has already rejected categorically as unlike himself and, therefore, to exert no lasting effects other than, perhaps, to reaffirm the correctness of his self conception. On matters of less consequence, he is more amenable to change and innovation, because he has not committed himself against such a broad range of discrepant values (Sherif, Sherif, & Nebergall 1965; Sherif, Sherif, & Kent 1967).

By studying the relative importance of a person's reference groups, their values, and his position within them, the degrees of a person's ego involvement in various respects may eventually become more amenable to research.

MUZAHER SHERIF

[Directly related are the entries IDENTITY, PSYCHOSOCIAL and PERSONALITY, article on PERSONALITY DEVELOPMENT. Other relevant material may be found in ATTITUDES; INFANCY; INTERACTION; LIFE CYCLE; PERSONALITY, article on THE FIELD; REFERENCE GROUPS; SOCIALIZATION.]

#### BIBLIOGRAPHY

- ALLPORT, GORDON W. 1943 *The Ego in Contemporary Psychology*. *Psychological Review* 50:451-478.
- ALLPORT, GORDON W. 1961 *Pattern and Growth in Personality*. New York: Holt.
- AUSUBEL, DAVID P. 1952 *Ego Development and the Personality Disorders*. New York: Grune & Stratton.
- BALDWIN, JAMES M. (1895) 1906 *Mental Development in the Child and the Race: Methods and Processes*. 3d ed., rev. New York and London: Macmillan.
- BUCENTAL, JAMES F. T.; and JEDIN, SEYMOUR L. 1950 Investigations into the "Self-concept." I. The W-A-Y Technique. *Journal of Personality* 18:483-498.
- CAMPBELL, ERNEST Q.; and PETTIGREW, THOMAS F. 1959 Racial and Moral Crisis: The Role of the Little Rock Ministers. *American Journal of Sociology* 64:509-516.
- CENTERS, RICHARD (1947) 1952 *The American Class Structure: A Psychological Analysis*. Pages 299-311 in *Society for the Psychological Study of Social Issues, Readings in Social Psychology*. Rev. ed. Edited by Guy E. Swanson et al. New York: Holt.
- CHAPMAN, DWIGHT W.; and VOLKMAN, JOHN A. 1939 A Social Determinant of the Level of Aspiration. *Journal of Abnormal and Social Psychology* 34:225-238.
- CHFIN, ISIDOR 1944 The Awareness of Self and the Structure of the Ego. *Psychological Review* 51:304-314.
- CLARK, KENNETH B.; and CLARK, MAMIE P. 1947 Racial Identification and Preference in Negro Children. Pages 169-178 in *Society for the Psychological Study of Social Issues, Readings in Social Psychology*. Edited by Theodore M. Newcomb and Eugene L. Hartley. New York: Holt.
- COOLEY, CHARLES H. (1902) 1956 *Human Nature and the Social Order*. Rev. ed. In Charles H. Cooley. *Two Major Works: Social Organization and Human Nature and the Social Order*. Glencoe, Ill.: Free Press. → Each title reprinted with individual title page and pagination. A separate paperback edition was published in 1964 by Schocken.
- FISHER, SEYMOUR; and CLEVELAND, SIDNEY E. 1958 *Body Image and Personality*. Princeton, N.J.: Van Nostrand.
- GOFFMAN, ERVING (1956) 1959 *The Presentation of Self in Everyday Life*. Garden City, N.Y.: Doubleday.
- HALL, CALVIN S.; and LINDZEY, GARDNER 1957 *Theories of Personality*. New York: Wiley; London: Chapman.
- HILGARD, ERNEST R. 1949 Human Motives and the Concept of the Self. *American Psychologist* 4:374-382.
- HOROWITZ, EUGENE L. 1944 Race Attitudes. Pages 139-247 in Otto Klineberg (editor), *Characteristics of the American Negro*. New York: Harper.
- JAMES, WILLIAM (1890) 1962 *The Principles of Psychology*. 2 vols. New York: Smith.
- KOCH, SIGMUND 1956 Behavior as "Intrinsically" Regulated: Work Notes Toward a Pre-theory of Phenomena Called "Motivational." Volume 4, pages 42-130 in *Nebraska Symposium on Motivation*. Edited by Marshall R. Jones. Lincoln: Univ. of Nebraska Press.
- KRECH, DAVID; CRUTCHFIELD, RICHARD S.; and BELLACHEY, EGERTON L. 1962 *Individual in Society. A Textbook of Social Psychology*. New York: McGraw-Hill. → First published in 1948 as *Theory and Problems of Social Psychology*, by Krech and Crutchfield.
- KUHN, MANFORD H.; and MCPARTLAND, THOMAS S. 1954 An Empirical Investigation of Self-attitudes. *American Sociological Review* 19:68-76.
- LECKY, PRESCOTT (1945) 1951 *Self-consistency: A Theory of Personality*. Rev. ed. New York: Island Press.
- LEWIS, MORRIS M. 1963 *Language, Thought and Personality in Infancy and Childhood*. New York: Basic Books.

- LINDESMITH, ALFRED; and STRAUSS, ANSELM 1949 *Social Psychology*. New York: Dryden.
- LURIA, ALEKSANDR R. 1961 *The Role of Speech in the Regulation of Normal and Abnormal Behavior*. New York: Liveright.
- MEAD, GEORGE H. 1913 *The Social Self*. *Journal of Philosophy* 10:374-380.
- MILLER, GEORGE A. 1962 *Psychology: The Science of Mental Life*. New York: Harper.
- MOUSTAKAS, CLARK (editor) 1956 *The Self*. New York: Harper.
- MURPHY, GARDNER 1947 *Personality: A Biosocial Approach to Origins and Structure*. New York: Harper.
- NEWCOMB, THEODORE M. 1950 *Social Psychology*. New York: Dryden.
- NEWCOMB, THEODORE M.; TURNER, RALPH H.; and CONVERSE, PHILIP E. 1965 *Social Psychology: The Study of Human Interaction*. New York: Holt.
- PIAGET, JEAN (1932) 1948 *The Moral Judgment of the Child*. Glencoe, Ill.: Free Press. → First published in French.
- PRADO, WILLIAM M. 1958 Appraisal of Performance as a Function of the Relative Ego-involvement of Children and Adolescents. Ph.D. dissertation, Univ. of Oklahoma. → Summarized in Sherif & Sherif (1964).
- RADCLIFFE-BROWN, A. R. (1922) 1948 *The Andaman Islanders*. Glencoe, Ill.: Free Press.
- REEDER, LEO G.; DONOHUE, GEORGE A.; and BIBLARZ, ARTURO 1962 Conceptions of Self and Others. Pages 69-78 in Bartlett Stoodley (editor), *Society and Self*. New York: Free Press.
- RIVERS, W. H. R. 1914 *The History of Melanesian Society*. 2 vols. Publication No. 1 of the Percy Sladen Trust Expedition to Melanesia. Cambridge Univ. Press.
- ROSENBERG, MORRIS 1965 *Society and the Adolescent Self-image*. Princeton Univ. Press.
- SCHNEIRLA, T. C. 1946 Problems in the Biopsychology of Social Organization. *Journal of Abnormal and Social Psychology* 41:385-402.
- SCHNEIRLA, T. C. 1951 The "Levels" Concept in the Study of Social Organization in Animals. Pages 83-120 in Conference on Social Psychology at Crossroads, University of Oklahoma, 1950, *Social Psychology at the Crossroads*. Edited by John H. Rohrer and Muzafer Sherif. New York: Harper.
- SCOTT, JOHN P. 1953 Implications of Infra-human Social Behavior for Problems of Human Relations. Pages 33-73 in Conference in Social Psychology, University of Oklahoma, 1952, *Group Relations at the Crossroads*. Edited by Muzafer Sherif and M. O. Wilson. New York: Harper.
- SECORD, PAUL F.; and BACKMAN, CARL W. 1964 *Social Psychology*. New York: McGraw-Hill.
- SEWARD, GEORGENE 1946 *Sex and the Social Order*. New York: McGraw-Hill.
- SHERIF, CAROLYN W.; SHERIF, MUZAHER; and NEBERGALL, R. E. 1965 *Attitude and Attitude Change: The Social Judgment-Involvement Approach*. Philadelphia: Saunders.
- SHERIF, MUZAHER 1962 The Self and Reference Groups: Meeting Ground of Individual and Group Approaches. New York Academy of Sciences, *Annals* 96:797-813.
- SHERIF, MUZAHER; and CANTRIL, HADLEY 1947 *The Psychology of Ego-involvements, Social Attitudes and Identifications*. New York: Wiley; London: Chapman & Hall.
- SHERIF, MUZAHER; and HARVEY, O. J. 1952 A Study in Ego Functioning: Elimination of Stable Anchorages in Individual and Group Situations. *Sociometry* 15: 272-305.
- SHERIF, MUZAHER; and SHERIF, CAROLYN W. 1956 *An Outline of Social Psychology*. Rev. ed. New York: Harper. → First published in 1948 with Muzafer Sherif as the sole author.
- SHERIF, MUZAHER; and SHERIF, CAROLYN W. 1964 *Reference Groups: Exploration Into Conformity and Deviation of Adolescents*. New York: Harper.
- SHERIF, MUZAHER; SHERIF, CAROLYN W.; and KENT, DONALD 1967 *Attitude, Ego-involvement and Change*. New York: Wiley.
- SHOEN, EDWARD J. JR. 1962 Behavioral Aspects of the Self. New York Academy of Sciences, *Annals* 96:765-773.
- SIEGEL, ALBERTA E.; and SIEGEL, SIDNEY 1957 Reference Groups, Membership Groups, and Attitude Change. *Journal of Abnormal and Social Psychology* 55:360-364.
- SMITH, M. BREWSTER; BRUNER, JEROME S.; and WHITE, R. W. 1956 *Opinions and Personality*. New York: Wiley.
- SNYGG, DONALD; and COMBS, ARTHUR W. 1949 *Individual Behavior*. New York: Harper.
- SYMONDS, PERCIVAL M. 1951 *The Ego and the Self*. New York: Appleton.
- WALLON, HENRI 1934 *Les origines du caractère chez l'enfant*. Paris: Boivin.
- WYLIE, RUTH C. 1961 *The Self Concept: A Critical Survey of Pertinent Research Literature*. Lincoln: Univ. of Nebraska Press.

## SELF DETERMINATION, NATIONAL See INTERNATIONAL POLITICS; NATION; NATIONALISM.

## SELIGMAN, C. G.

Charles Gabriel Seligman (1873-1940) was one of that remarkable group of young biological scientists (among them, W. H. R. Rivers, C. S. Myers, and William McDougall) who, under A. C. Haddon's inspiration, established scientific field work as the basis of professional anthropology in Great Britain during the first quarter of the twentieth century. Seligman's research broke ground for Malinowski's field work in Melanesia between 1914 and 1918 and for Evans-Pritchard's work in the Sudan twenty years later.

Seligman drifted into anthropology by chance. Originally trained as a physician, he had embarked on research in medicine when, in 1898, he learned of Haddon's plans for the Cambridge Expedition to the Torres Strait and volunteered to join it. At first Seligman confined himself to medical investigations, but his colleagues soon drew him into their anthropometric and ethnological studies. On



returning to England in 1899, he resumed his medical research until, in 1903, he met a wealthy American businessman, Major Cooke Daniels, and persuaded him to finance the expedition to New Guinea that resulted, *inter alia*, in *The Melaneians of British New Guinea* (1910). It was only after this expedition that he finally committed himself to anthropology by accepting the first lectureship in ethnology (in 1913 it became the chair of ethnology) established at the University of London.

In 1904 he married Brenda Z. Salaman, who collaborated in all his later field research and in many of his major publications. Her influence on their joint work was important: while Seligman himself had, for example, little interest in the minutiae of kinship custom, his wife, under the influence and instruction of Rivers, became adept in this branch of anthropology and took the responsibility for it. The Seligmans carried out a series of officially sponsored expeditions, to Ceylon in 1907–1908 to study the Veddās (see 1911) and their first expedition to the Sudan in 1909–1910. They made two further trips to the Sudan in 1911–1912 and 1921–1922.

In the intervals between these expeditions, while preparing their results for publication, Seligman continued his pathology researches and received for them, in 1911, the coveted distinction of a fellowship of the Royal College of Physicians, as well as other medical honors. In 1919 he also became a fellow of the Royal Society.

Like Haddon and Tylor, whose conception of a "science of culture" he admired, Seligman regarded anthropology as embracing every aspect of the life of the simpler societies (1926). He approached field work in the spirit of a Darwinian natural historian, considering a living society in terms of its geographical and historical setting as well as in terms of its racial composition, cultural make-up, and psychological foundations. Gifted also with a sensitive aesthetic judgment, Seligman built up a choice private collection of early Chinese art, notably of ceramic and bronze ware, and was thus drawn into research on Chinese history and civilization, a subject that was peculiarly congenial to his somewhat reserved temperament.

During World War I he joined the Royal Army Medical Corps and was, with Rivers, assigned to the staff of a hospital for victims of shell shock. Clinical experience convinced him of the general soundness of Freudian theory, and in later essays he applied psychoanalytical and related theories to the problems of ethnology.

**Descriptive ethnology.** Two features are characteristic of Seligman's anthropological and ethnological contributions: they are exploratory rather than definitive in aim, and they concentrate on factual data as opposed to theoretical interpretations and generalizations. Seligman's expedition to New Guinea set the pattern for his later ethnological work. Its aim was to establish the distinctive racial, social, and cultural features of the peoples of this region, with a view to their ethnological classification. All available sources were utilized, including the works of travelers, government reports, and notes and memoranda provided by long-resident missionaries and administrators, but the chief emphasis was on information obtained from native informants and eyewitness description.

Considering that the field work took less than a year and was carried out either in pidgin English or with the aid of often inadequate interpreters, the results are impressive. Malinowski (1922) paid tribute to the excellence of Seligman's observations. The data were for the most part new to anthropology. Where they were doubtful or defective, this was scrupulously indicated. Though Seligman made no attempt to investigate any subject in the depth achieved by the next generation of ethnographers, he did deal quite exhaustively with some topics (e.g., totemism), and he did not fail to draw penetrating conclusions, as in his elucidation of the powers and authority of chiefs.

Seligman's 1910 book remains a "basic and systematic work" on Melanesia (Elkin 1953, p. 47). It covers, in varying detail, every significant facet of tribal life: the geographical setting, racial composition, kinship and family structure, chieftainship and rank, property, trade, warfare, morals, magic and religion, material culture, feasts and ceremonies. It contains an important and comprehensive account (contributed by F. R. Barton) of the *kula*-type trading voyages of the Motu and an outline of Northern Massim culture on which Malinowski, and subsequently Reo Fortune, built. [See OCEANIAN SOCIETY.]

The field work that resulted in the two joint works entitled *The Veddās* (1911) and *Pagan Tribes of the Nilotic Sudan* (1932) followed the same procedures as those used by Seligman in New Guinea. *The Veddās* represents a valiant attempt to salvage what could still be found out about the culture and mode of life of the remnants of the aborigines of Ceylon, and it remains a standard work. The results of the Sudan venture, however, turned out to be more important. The three

Sudan expeditions, prepared for with exemplary thoroughness and supported by both the Sudanese government and British learned societies, were pioneer undertakings in every sense. In this huge territory, still barely pacified, living and traveling conditions were arduous. Communication with informants, in Arabic, was difficult; the Seligmans had only a working knowledge of the language and the southern Sudanese often no more than a smattering of it. Nevertheless, the Seligmans opened up an area until then ethnologically uncharted.

The works they published between 1911 and 1932 about this research made ethnological history. In particular, it was Seligman's description of the cult of Nyakang and the divine kingship of Shilluk (see 1911) that, as Evans-Pritchard commented, "brought . . . [divine kingship] to the notice of Sir James Frazer and into the main stream of ethnological theory" (1948, p. 1). *Pagan Tribes* summed up the results of the field research of the Seligmans and other investigators during the twenty years following the first expedition.

Seligman espoused the theory that successive waves of pastoral Hamites had, in the distant past, spread via the Horn of Africa westward and southward, mingling with the autochthonous Negro population to produce the Nilotes (Shilluk, Dinka, Nuer, etc.) and Nilo-Hamites (Bari, Lotuko, etc.) and the chain of peoples in the Great Lakes region stretching through the Hima elements (e.g., in Ankole) to the Masai, all distinguished by their greater or lesser adherence to pastoralism (1913). Conjectural as the theory of Hamitic migration may be, the ethnic classification remains valid, as its continued use (e.g., in the Greenberg language classification) testifies. The theory became a cornerstone of Seligman's synoptic review of the peoples and cultures of Africa (1930). These are divided into six major ethnic clusters—the Bushmen and Hottentots, the true Negroes of west Africa, the Hamites (eastern and northern), the Nilo-Hamites and Nilotes, the Bantu, and the Semites. It is a singular feat of compression and still the best over-all introduction to African ethnology. [See AFRICAN SOCIETY, article on SUBSAHARAN AFRICA.]

**Historical ethnology.** Seligman's concern with historical ethnology grew naturally out of his descriptive studies. It was also a response to the challenge of the extreme diffusionist theories advocated by Elliott Smith and Rivers. Seligman flatly rejected these theories, contrasting them with genuinely historical reconstruction based on geo-

graphical connections and evidence derived from material culture and institutional relationships (1927). Thus, when he himself carefully considered the evidence for the supposed Egyptian origin of divine kingship in Africa, he came to the conclusion that its most characteristic features go back to an "old and widespread Hamitic belief" to which some specifically Egyptian elements only later became attached (1934). [See KINGSHIP.]

It was in the pursuit of these and other historical and distributional associations that Seligman undertook his investigations into prehistory (e.g., of Egypt and the Sudan) and his studies of artifacts and *objets d'art*. Coupled with his passion for Chinese art and civilization, these studies culminated in his joint paper, with H. C. Beck, on Chinese glass (1938). Here Seligman sought to clarify the routes and modes of transmission by which various products and technical skills reached China from western Europe even before the second century B.C.

**Physical anthropology.** The investigations in physical anthropology that formed for Seligman the necessary basis for the organization of cultural and historical data now appear relatively crude. Without such modern techniques as genetic analysis of blood groups, he obtained anthropometric and anthroposcopic data that have only limited validity.

**Cross-cultural psychology.** Seligman's psychological investigations began with the application of sensory discrimination tests to the Veddas; the results were discouraging. Later he became convinced that the psychology of the unconscious was the key to many fundamental problems of anthropology, and in his first presidential address to the Royal Anthropological Institute he stressed "the borderland where social anthropology, psychology, and genetics meet in common biological kinship" (1924, p. 13). (He reverted to this theme in his Huxley Memorial Lecture in 1932). He enlisted the aid of field ethnologists, psychoanalysts, orientalist, and a variety of other specialists, as well as laymen, in the pursuit of these inquiries. Malinowski was one of the many who owed his introduction to psychoanalysis to Seligman (Malinowski 1927, preface).

Guided by Freud's theory, Seligman produced evidence from many different cultures for the universality of such "type" dreams as those involving the loss of teeth, flying, and raw meat. He found that conventional interpretations of such dreams often implicitly recognized their symbolic nature. He inferred that there is in the unconscious of all



peoples "a proved common store on which fantasy may draw" in the construction of myth and ritual beliefs and even of technical processes (1924, p. 41). His interpretation of magical rites was that they are "symbolic dramatizations" that drain away unconscious anxiety and that are thus comparable to neurotic symptoms. Taking a different perspective, he assembled cross-cultural evidence to test the universality of the stages of individual psychosexual development postulated in Freudian theory. Emphasizing the scanty and conflicting nature of the ethnological data then available, he concluded that they tended to confirm rather than to refute (as Malinowski maintained) the Freudian hypothesis (1932). Also impressed with Jung's extrovert-introvert dichotomy of temperamental types, he applied it in tentative studies of art styles, of the national character of the Japanese, and of psychosis in different cultural settings (1931). He concluded that psychosis was rare in stable primitive societies but tended to emerge as a result of conflicts engendered by contact with European culture, conflicts he connected also with the then incipient nativistic movements now called cargo cults.

These studies claimed no more than to demonstrate the significance of psychoanalysis for anthropology at a time when there was a strong reaction in British anthropological circles against such an application of psychoanalytic theories. [See CULTURE AND PERSONALITY.]

Just as Seligman rejected all-embracing diffusionist explanations of ethnological connections, so also he was chary of evolutionist constructions of the kind favored by Frazer and L. H. Morgan. With his strong biological bias, he saw living mankind as a single species with common and uniform mental and physical capacities and dispositions but distributed over the globe in racially diverse and culturally varied communities in different phases of historical development. Though Malinowski's functionalism grew up directly under his eyes, and not without his sympathy, he never aligned himself with this movement; he was too much of an empiricist. The task of scientific ethnology, as he saw it, was to approach particular areas and particular problems by as many methods and from as many angles as could serve to reveal useful data and throw light on them.

MEYER FORTES

[Other relevant material may be found in the biographies of FRAZER; HADDON; MALINOWSKI; RIVERS; TYLOR.]

#### WORKS BY SELIGMAN

- 1910 *The Melanesians of British New Guinea*. Cambridge Univ. Press.
- 1911 SELIGMAN, C. G.; and SELIGMAN, BRENDA Z. *The Veddas*. Cambridge Univ. Press.
- 1913 Some Aspects of the Hamitic Problem in the Anglo-Egyptian Sudan. *Journal of the Royal Anthropological Institute of Great Britain and Ireland* 43:593-705.
- 1918 SELIGMAN, C. G.; and SELIGMAN, BRENDA Z. *The Kababish: A Sudan Arab Tribe*. *Harvard African Studies* 2:105-185.
- 1924 Anthropology and Psychology: A Study of Some Points of Contact. *Journal of the Royal Anthropological Institute of Great Britain and Ireland* 54:13-46.
- 1926 Anthropology. Volume 1, pages 131-143 in *Encyclopædia Britannica*, 13th ed. Chicago: Benton.
- 1927 [A Book Review of] *Psychology and Ethnology*, by W. H. R. Rivers. *Nature* 120:685-687.
- 1928 The Unconscious in Relation to Anthropology. *British Journal of Psychology* 18, part 4:373-387.
- 1929 Temperament, Conflict and Psychosis in a Stone-age Population. *British Journal of Medical Psychology* 9, part 3:187-202.
- (1930) 1957 *Races of Africa*. 3d ed. Oxford Univ. Press.
- 1931 Japanese Temperament and Character. *Japan Society of London, Transactions and Proceedings* 28:124-138.
- 1932 Anthropological Perspective and Psychological Theory. The Huxley Memorial Lecture. *Journal of the Royal Anthropological Institute of Great Britain and Ireland* 62:193-228.
- 1932 SELIGMAN, C. G.; and SELIGMAN, BRENDA Z. *Pagan Tribes of the Nilotic Sudan*. London: Routledge.
- 1934 *Egypt and Negro Africa: A Study in Divine Kingship*. The Frazer Lecture for 1933. London: Routledge.
- 1938 SELIGMAN, C. G.; and BECK, H. C. *Far Eastern Glass: Some Western Origins*. Stockholm, Östasiatiska samlingarna, Bulletin no. 10:1-64.

#### SUPPLEMENTARY BIBLIOGRAPHY

- ELKIN, ADOLPHUS P. 1953 *Social Anthropology in Melanesia: A Review of Research*. Oxford Univ. Press.
- EVANS-PRITCHARD, E. E. 1948 *The Divine Kingship of the Shilluk of the Nilotic Sudan*. Cambridge Univ. Press.
- FORTES, MEYER 1941 Charles Gabriel Seligman. *Man* 41, no. 1:1-6.
- HADDON, A. C. 1934 Appreciation. Pages 1-4 in *Essays Presented to C. G. Seligman*. Edited by E. E. Evans-Pritchard et al. London: Routledge.
- HUNTINGFORD, GEORGE W. B. 1953 *The Northern Nilotes*. Ethnographic Survey of Africa, East Central Africa, Part 6. London: International Africa Institute.
- MALINOWSKI, BRONISLAW (1922) 1960 *Argonauts of the Western Pacific: An Account of Native Enterprise and Adventure in the Archipelagoes of Melanesian New Guinea*. London School of Economics and Political Science, Studies, No. 65. London: Routledge; New York: Dutton. → A paperback edition was published in 1961 by Dutton.
- MALINOWSKI, BRONISLAW (1927) 1953 *Sex and Repression in Savage Society*. London: Routledge; New York: Harcourt. → A paperback edition was published in 1955 by Meridian.
- MYERS, C. S. 1941 Charles Gabriel Seligman. *Royal Society of London, Obituary Notices of Fellows* 3:627-646. → Includes a bibliography of Seligman's works.

## SELIGMAN, EDWIN R. A.

Edwin Robert Anderson Seligman (1861–1939), American economist, was editor-in-chief of the *Encyclopaedia of the Social Sciences* from 1927 to 1935 and McVickar professor of political economy at Columbia University from 1904 to 1931. He was born in New York City, where his father was a banker prominent in national and international circles. Until the age of 11, Seligman was educated at home, his tutor being Horatio Alger, Jr. From the Columbia Grammar School Seligman entered Columbia College, shortly after his fourteenth birthday, and received his A.B. in 1879. He became proficient in the German, French, Italian, Spanish, Russian, and Dutch languages. After attending the universities at Berlin, Heidelberg, and Paris, in 1882 he began graduate studies in economics at Columbia and, "to have a second arrow in the quiver," also studied law. In 1884 he received both an M.A. and an LL.B., and in 1885, his Ph.D. Also in 1885 Seligman was admitted to the New York bar and was appointed lecturer in economics in the newly formed faculty of political science at Columbia University. Three years later he became adjunct professor of political economy; in 1891 he was appointed professor of political economy and finance, a post he held until 1931. He served for many years as chairman of the department of economics. In the 1920s Seligman received honorary doctoral degrees from the universities of Paris and Heidelberg.

Seligman's immense energy, wide-ranging interests, and sense of social responsibility led him to participate in the work of academic, governmental, and civic organizations and committees, especially those promoting social reform. He was a member of the Social Reform Club of New York City, the Educational Alliance, the Political Economy Club, Greenwich House (a settlement house), and the Society for Ethical Culture and served as president of the last two organizations and as chairman of the Bureau of Municipal Research. He was secretary of the Committee of Fifteen, whose 1902 report, *The Social Evil*, revealed the economic links of corruption between prostitution and the New York City police.

Seligman was a leader in the founding of the American Economic Association (AEA) and served as its president from 1902 to 1904. He also served a term as president of the National Tax Association, from 1913 to 1915. He helped to establish both the City Club of New York, which has worked for the improvement of civic conditions since its founding in 1892, and the American Association

of University Professors (AAUP), in 1915; he was AAUP president from 1919 to 1920. Seligman was an effective champion of academic freedom. He was the chairman of the AEA committee that wrote the first report on this subject in the United States, the occasion being the dismissal of Edward A. Ross from Stanford [see the biography of Ross], and he was also chairman of the AAUP committee that in 1915 wrote the fundamental *Report on Academic Freedom*. In the field of his specialty, public finance, Seligman was also active: he was a member of special tax commissions of New York State in 1906 and 1930; chairman of the Mayor's Commission on Taxation and Finance, New York City, 1914–1916; expert adviser to the New York State Joint Legislative Tax Committee, 1919–1922; member, with Bruins, Einaudi, and Stamp, of the League of Nations Committee on Economics and Finance, 1922–1923, which produced the report *Double Taxation*; and, in 1931, consultant to the government of Cuba.

When the study of political economy at Columbia became more specialized, Seligman accepted responsibility for the then neglected field of public finance, but he also developed a continuing interest in the history of economic doctrine. This interest led to his giving a course on this subject and to his famous collection of books, pamphlets, and other items on economics, now in the Columbia University library. He was one of the small group of scholars that established economics as a recognized discipline in the United States.

Seligman's intellectual influence was wide, several of his books being translated into more than one of the leading languages. It has extended to the present, largely through his contributions to the history of economic thought, reflecting the stimulus he received at Paris, and through the clarity and appropriateness of his conceptual and categorizing contributions. The former achievement is illustrated by his article "On Some Neglected British Economists" (first published in 1903; see [1886–1925] 1925); the latter, by the introductory chapter of *The Shifting and Incidence of Taxation* (1892), "The Classification of Public Revenues" (see 1895), and his analysis of ability-to-pay and faculty concepts in taxation (see 1928). More than one present-day treatise, particularly in Europe, refers frequently to Seligman's conceptual and historical contributions.

Although Seligman did not entirely escape error in attempting to apply the newly developing price theory to the substantive part of *Shifting and Incidence*, he did enliven the doctrinal atmosphere with his vigorous approach and equally vigorous coun-



terthrusts when challenged. His exchanges with Francis Y. Edgeworth were enlightening. Seligman's treatise was a pioneering achievement in its treatment of the history of doctrine and in its arrangement and development of a difficult subject.

Seligman's writings on taxation, quite influential in their day, are less cited now, either because the paths he opened up were later explored further by so many, or because the terminology he originated soon passed into common use, or because the reforms he advocated were adopted. Among his useful contributions were his *Progressive Taxation in Theory and Practice* (1894), *The Income Tax* (1911), and his writings on the property tax, all of which also embody his desire, arising from the contrast between his American and European training, to build a bridge between the deductive and the historical-comparative schools, with emphasis on the latter, which he felt had been neglected in the United States.

Seligman also wrote extensively outside public finance. Some of these contributions were widely accepted for many years, especially *The Economic Interpretation of History* (1902), which proved influential in developing analysis in this area. Among his interests in applied economics were problems of railroads, unemployment, the fur and jewelry trades, installment selling, agriculture, and resale price maintenance.

It was this catholicity of interest, together with an exceptional erudition and a broad acquaintance with scholars of many countries, that enabled Seligman, toward the close of his career, to make one of his most notable contributions, as a chief promoter of, fund raiser for, and editor-in-chief of the English-language *Encyclopaedia of the Social Sciences*, to which he also contributed articles and biographies.

Seligman, formal in the lecture room, was a man of warm temperament, with an intense personal interest in the material and intellectual welfare of his students. Through them and through his own writings, he opened up the field of public finance in the United States both to scholarship and to measures of reform.

CARL S. SHOUP

[See also INSTALLMENT CREDIT; TAXATION.]

#### WORKS BY SELIGMAN

- (1886-1925) 1925 *Essays in Economics*. New York: Macmillan.  
 1887 *Two Chapters on the Mediaeval Guilds of England*. Baltimore: American Economic Association.  
 (1892) 1927 *The Shifting and Incidence of Taxation*. 5th ed., rev. New York: Columbia Univ. Press.

- (1894) 1908 *Progressive Taxation in Theory and Practice*. 2d ed., rev. & enl. Princeton, N.J.: American Economic Association.  
 (1895) 1928 *Essays in Taxation*. 10th ed., rev. New York: Macmillan.  
 (1900-1925) 1925 *Studies in Public Finance*. New York: Macmillan.  
 (1902) 1934 *The Economic Interpretation of History*. 2d ed., rev. New York: Columbia Univ. Press. → A paperback edition was published in 1961.  
 (1905) 1929 *The Principles of Economics, With Special Reference to American Conditions*. 12th ed. New York: Longmans.  
 (1911) 1914 *The Income Tax: A Study of the History, Theory, and Practice of Income Taxation at Home and Abroad*. 2d ed., rev. & enl. New York: Macmillan.  
 1926 *Social Theory of Fiscal Science*. *Political Science Quarterly* 41:193-218, 354-383  
 1927 *The Economics of Instalment Selling: A Study in Consumers' Credit, With Special Reference to the Automobile*. 2 vols. New York and London: Harper.  
 1928 *Double Taxation and International Fiscal Cooperation*. New York: Macmillan.  
 1929a *Autobiography*. Volume 2, pages 117-160 in Felix Meiner (editor), *Die Volkswirtschaftslehre der Gegenwart in Selbstdarstellungen*. Leipzig: Meiner.  
 1929b *The Economics of Farm Relief: A Survey of the Agricultural Problem*. New York: Columbia Univ. Press.  
 1930 *What Are the Social Sciences?* Volume 1, pages 3-7 in *Encyclopaedia of the Social Sciences*. New York: Macmillan.  
 1932 *Income Tax*. Volume 7, pages 626-639 in *Encyclopaedia of the Social Sciences*. New York: Macmillan.  
 1932 SELIGMAN, EDWIN R. A.; and LOVE, ROBERT A. *Price Cutting and Price Maintenance: A Study in Economics*. New York and London: Harper.  
 1934 *Public Finance*. Volume 12, pages 637-646 in *Encyclopaedia of the Social Sciences*. New York: Macmillan.

#### SUPPLEMENTARY BIBLIOGRAPHY

- [The Bibliography of] Edwin Robert Anderson Seligman. 1931 Pages 25-43 in Columbia University, Faculty of Political Science, *A Bibliography . . . 1880-1930*. New York: Columbia Univ. Press.  
 DORFMAN, JOSEPH (editor) 1941 *Seligman Correspondence*. *Political Science Quarterly* 56:107-124, 270-286, 392-419, 573-599.  
 DORFMAN, JOSEPH 1946-1959 *The Economic Mind in American Civilization*. 5 vols. New York: Viking.  
 DORFMAN, JOSEPH 1955 *The Department of Economics*. Pages 161-206 in Ralph G. Hoxie et al., *A History of the Faculty of Political Science, Columbia University*. New York: Columbia Univ. Press.  
 Edwin Robert Anderson Seligman, 1861-1939: *Addresses Delivered at the Memorial Meeting Held on December the Thirteenth, 1939*. 1942 Stamford, Conn.: Overbrook.

#### SEMANTICS AND SEMIOTICS

Semiotics is the study of sign phenomena. Specialized research into natural human language—the semiotic phenomenon par excellence—constitutes linguistics; within linguistics, semantics is

concerned with the conveyance of meaning by the grammatical and lexical devices of a language. According to the theoretical, descriptive, and historical slants of linguistic investigation, semantic problems respectively assume a general, synchronic, or diachronic character.

The self-evident systematicity of grammatical phenomena has always been conducive to their relatively reliable semantic analysis. When it comes to the looser domain of vocabulary, however, the obscurity of the underlying structure quickly embroils semantic analysis in some of the more inconclusive epistemological controversies of social science. While vast supplies of raw semantic data repose in dictionaries of various languages, there is no consensus among linguists on a coherent theory in accordance with which this material can be analyzed and compared for purposes of generalization.

**Definition and components of sign.** The key term in semiotics is "sign." By all accounts, a sign must consist of at least (1) a perceptible impact on one of the senses of the "receiver" and (2) a content that is signified. First distinguished by the Stoic philosophers, then recognized in European medieval literature as *signans* and *signatum*, these minimal components of the sign have reappeared in modern theory under such labels as *signifiant* and *signifié* (Ferdinand de Saussure) or "sign vehicle" and "designatum" (Charles W. Morris). (Under appropriate contextual conditions, the absence of a *signans* itself functions as a "zero" sign vehicle. It is often utilized in language with great effectiveness.) A particular occurrence of a sign is customarily called a "token," while the class of all occurrences of the sign is called the "type." According to this usage, the total number of words in a text is the number of word tokens, whereas the number of different words in the text is the number of word types.

Signs may be classified by the sense on which the sign vehicle impinges: they can be visual, auditory, etc. A sign may then have equivalent sign vehicles in the same or another modality. Full equivalence is illustrated by a plain transposition cipher (based on a novel utilization of the same graphic system) or by a light blinking and a buzzer sounding in response to the same Morse code impulses. The correlation, on the other hand, between written and spoken language is far from one-to-one.

**Designation and denotation.** A crucial issue for most students is the distinction between what particular sign tokens denote, or refer to, and the

constant capacity of the sign type to denote, or refer. The former relation has been treated under the heading of "denotation," or reference; the latter has been called "designation," or meaning (in the narrow sense). Thus, the designatum (2a) of a sign type—what the sign designates—can in theory be specified by (2b) the conditions under which, or the "objects" with reference to which, the tokens of the sign would be said to denote in a truthful way. Every sign by definition designates something, but some signs have no real things, or "denotata," to refer to (e.g., *unicorn*), and every sign may on occasion be used without denoting (e.g., *milk* when no actual milk is being referred to) or without denoting truly (e.g., *milk* used with reference to water). A perceptible discrepancy between what a sign type designates (2a) and the denotatum of one of its tokens (2b) seems to be involved in such effects as metaphor and irony, as well as in perversions of communication (lying).

If the designatum is thought of as the class of denotata for which the sign vehicle is a name, the specification of the designatum can correspondingly be visualized as extensional (by enumeration of the members of the class, i.e., the denotata) or as intensional (by stating the properties shared by all members of the class, and only by them). The ability of living beings to form intensional class concepts—an ability which has not so far been mechanically simulated or abstractly reconstructed—reaches the most extraordinary proportions in man. There is evidence that concept formation proceeds with particular efficiency if there are signs present for which the concepts can become the designata.

The theoretical distinction between (2a) designata and (2b) denotata is also essential for accommodating the fact that something can be a true denotatum of more than one simple sign (e.g., *car* and *vehicle*) or of various compound signs (e.g., *female sheep* = *ewe*).

**"Reductionist" approaches.** The minimal account of signs in terms of sign vehicles and designata has been variously enriched; the richer schemes have in turn been subjected to reduction along several lines. Two further factors of the sign, essential in some accounts, such as that of Charles S. Peirce, are (3) the interpreter, or the individual for whom the pairing of a sign vehicle and a designatum functions as a sign, and (4) the interpretant, corresponding to the interpreter's (perhaps unconscious) reaction to a sign. One way of explicating this reaction is to understand "interpretant" as the interpreter's private translation of a sign into other signs. Anticipating the discovery



of neurological correlates of "understanding" and similar psychological phenomena, some theorists have apparently identified the interpretant of a sign with its translation into the (still unknown) neurological code.

Related in spirit is the effort of experimental psychologists to specify the meaning of a sign in terms of attitudes taken toward it, emotions induced by it, or further signs automatically evoked by it—as displayed, for example, in a semantic differential (Osgood et al. 1957) or a free-association test. Thus, whereas the designatum of a sign was traditionally understood to be given by the verbal definition of the sign, behavioristic psychology has endeavored to bypass the designatum (2a) in favor of the effect of the sign (4), regarding the latter, if not as more accessible to study, then at least as the less objectionable construct. A kindred suspicion of constructs has moved skeptical empiricists—in eighth-century India as in twentieth-century England and America—to reject designata as needless entities and to account for the use of sign tokens (acts of "reference") merely in terms of the sign vehicle (1) and the actual denotatum or "referent" (2b).

Such impoverished theories may cover certain "straightforward" uses of natural language and may serve the rationalization of scientific discourse. But an account that fails to treat the designatum as a component of the sign, distinct from both its denotatum and its interpretant, cannot adequately deal with such phenomena as reasoning or humor, and is therefore incommensurate with the complexity and subtlety of human semiotic abilities.

In general it should be admitted that the denial of the *sui generis* nature of communicative activity and the reduction of sign phenomena to some more general kind of behavioral phenomenon have produced no marked success in either theory or research practice. A sign type is not always—perhaps only rarely—correlated with a class of specific stimuli or of overt responses. Nor is a sign token a counterfeit or avowed "substitute" for a "real" thing; it is, at most, a "representative" of the real things for purposes of communication qua communication, and in contrast to the case of other substitutes, the real thing will not do where a sign is required. When we reflect, furthermore, that the denotata of most sign tokens of language are neither "real" nor "things" nor even necessarily existent, it becomes evident that if sign tokens "stand for something," they do so in a way that differs from other modes of substitution.

**Symbolic and less-than-symbolic signs.** A sign with an intensional class for a designatum and without contiguity or similarity between its vehicle

and its denotata is called a symbol. Other, less fully developed forms of sign may be classified as indices, icons, names, signals, and symptoms.

A sign is said to be an index rather than a symbol insofar as its sign vehicle is contiguous with its denotatum, or is a physical sample of it (a swatch of cloth as a sign of the color or an onomatopoeic word as a sign for an animal sound). In the case of some indices the contiguity between the sign vehicle and the denotatum is suggestive rather than literal, as in a directional arrow which guides the viewer's eyes toward a target. In general, when a sign vehicle is paired with a denotatum as a matter of necessity, the involuntary, mechanical nature of this connection may be viewed as an abstract analogue of physical contiguity; automatic, nonarbitrary signs are hence said to be merely "symptomatic" rather than fully symbolic.

Natural languages have been observed to contain indexlike signs which conventionally direct the listener's attention to a time, a place, a participant in the discourse, or a part of the interlocutor's field of sight. These are studied under the heading of "deictics" or "shifters"; English examples are *now*, *here*, *I*, *this*.

When there is a geometric similarity between a sign vehicle and its denotata, the sign is said to be *iconic*. Such similarity would be exemplified by a system in which, let us say, large things are signified by long words, small things by short words, or in which plurality of denotata is signified by repetition of the sign vehicle. A realistic painting is a highly iconic sign; in human language the role of iconicity is marginal.

A theory of signs which conceives of a designatum as an intensional class must also allow for a more "stunted" type of sign which has an extensional class as its designatum. Such signs are generally called "(proper) names." The individuals, whether one or many, which are truly denoted by a proper name have no common property attributed to them except "answering" to that name.

For discourse about signs it is necessary to use signs for referring to signs. For this purpose a sign vehicle is commonly employed as an index of its own sign. In discourse about languages it has long been found useful to discriminate between the use of signs and the mention of signs. A sign employed as a name for itself is commonly said to be used "metalinguistically"; a specialized language for communication about another language (the "object language") is called a "meta-language."

**Overlapping semiotic functions.** In a particular communicative act a token may at the same

time function as a symbol and as a less-than-symbolic sign. Thus, a sign token may trigger, mechanically or conventionally, some action on the part of the receiver; it then functions as a "signal." In many languages, the imperative is a grammatical device for endowing a symbol with signal value, but a complex symbolic sign without special markers (e.g., *I am ready*) may function secondarily as a signal for some action. Similarly, an interjection of pain, to the extent that it has a coded form in the language (e.g., English *ouch* vs. German *au*), is conventionalized and hence symbolic; insofar, however, as it is uttered involuntarily, it is a symptomatic (indexical) sign. Every layman develops considerable skill in synchronizing the "symbolic" analysis of speech with a judgment of voice quality, tempo, and other involuntary aspects of the utterance as symptoms of the speaker's state. Specialists have learned to interpret in depth certain "covert," involuntary statistical properties of speech to which ordinary hearers may not respond even on an unconscious level.

Not only do symbolic functions overlap with symptomatic and signaling sign functions, but the sign may, primarily or secondarily, serve altogether noncommunicative functions as well. Superimposed upon linguistic utterances with symbolic value may be aesthetic or magical functions (poetry, incantations). Contrariwise, behavior patterns and artifacts intended for other primary purposes may acquire a signlike aspect: a garment, worn to provide warmth, may by its shape symbolize the wearer's acceptance of, or rebellion against, the conventions of society; an automobile, primarily a means of transport, may by the redness of its color symbolize its user's status as fire chief. The development of general and comparative semiotic research would seem to depend to a large extent on the inclusion of such "mixed" semiotic phenomena.

**Paradigmatic relations between signs.** Two or more signs each or all of which can occur in the same context are said to form a paradigmatic set. Membership in such a set helps to determine the identity of a sign, since the definition of its sign vehicle and its designatum may be formulated in terms of the discrete differences between them and the vehicles and designata of other signs in the same set. Students of language have capitalized on the paradigmatic nature of their material by organizing the description of sign vehicles and designata around those minimal distinctive differences of sound and meaning which contrast one item with another within the total system. In the investigation of vocabulary, this approach has led

to the concept of "word fields," i.e., semantically related groups of words, and has yielded a sizable literature since it was initiated in the 1930s by Jost Trier. However, the more populous and amorphous a paradigmatic set of elements, the less certain is the organization of their contrastive features. Hence word-field studies are beset by a strong streak of impressionism, exacerbated by the concentration of research on early stages of languages for which the benefit of native speakers' intuitions cannot be drawn upon.

Being different from each other is, of course, only the most general relation between signs in a paradigmatic set. More specific relations are determined by the conditions under which two signs are interchangeable: symmetrically in the case of perfect synonyms, asymmetrically in the case of superordinates and subordinates (e.g., *gun-rifle*), and at the cost of reversing the truth value of the message in the case of antonyms. The way in which these more specific relations organize a set of terms may be different in various languages. Thus, two languages may have terms for "cattle," the difference between them being that in one of the languages, but not in the other, "cattle" includes "sheep" as a subordinate.

Proceeding from their experience with folk classification in the field of kinship, anthropologists (particularly in the United States) have analyzed selected sectors of vocabulary in the form of taxonomies—systems in which all terms are governed by a subordinate-superordinate relation (Conklin 1962). It still remains to be shown whether this descriptive format is easily applicable to lexical domains less closely structured than those dealing with kinship, color, weather, illness, plants, and animals. It is clear, moreover, that studies of lexical systems oversimplify the problem unless they take full account of the omnipresent facts of polysemy, grammatical specialization, and phraseological specialization (discussed below). Meanwhile the introduction of certain nonsymmetrical operations to supplement the traditional algebra of classes promises to reduce some of the counterintuitive excesses of earlier sophisticated nomenclature analysis (Lounsbury 1964).

**Sign tokens in sequence.** The patterning of sequences in which sign tokens may be transmitted ranges between two extremes. At one pole lies the impossibility of predicting a sign's occurrence from a preceding sign; at the other extreme lie systems with a completely predictable sequence, such as the alternation of red and green at an automatic traffic signal. A sign system like human language is vastly more intricate than these models because of two groups of factors: the interde-









rugged opponent of tradition who became Senior's lifelong friend and exerted considerable influence on him. Senior entered Lincoln's Inn in 1813 to read for the bar; in 1817 he became a certified conveyancer and was called to the bar two years later. Although he earned his living by the practice of his profession, the law was by no means his only interest.

Senior's childhood and youth had been spent in his father's country parish, which meant that early in his life he had become familiar with the poverty and pauperization of agricultural laborers that had occurred during the Napoleonic wars. This experience made a deep impression on him and was undoubtedly one of the reasons for his enduring interest in economics. In 1823 he became a member of the Political Economy Club, to which he was introduced by James Mill; he was active there in discussions with his fellow economists. Shortly after, in 1825, he became the first holder of the Drummond professorship of political economy at Oxford. During his five-year term he published a number of important short works on economics, starting with an analysis of ambiguous economic terms that was published as an appendix to Whateley's *Elements of Logic* (Senior 1826). He published some of his lectures, on such subjects as methodology (1827), the theory of money and international trade (1830a), and population and wages (1829; 1830b). These lectures, together with the others given during his first term in the Drummond chair, were assembled and published as *An Outline of the Science of Political Economy* in 1836. As soon as they appeared, they established Senior's reputation as an economist.

### Role in Whig public policy

By 1830 Senior was well known to leading Whig politicians; he was invited by Lord Melbourne to investigate the laws relating to trade combinations, and Lord Howick suggested he examine the problems of Ireland. His memorandum on trade combinations was not published, but that on Ireland appeared in 1831 as *A Letter to Lord Howick: On a Legal Provision for the Irish Poor, Commutation of Tithes, and a Provision for the Irish Roman Catholic Clergy*. In this pamphlet he advocated the transfer of some of the property of the Anglican church to the Roman Catholic church in Ireland, and as a consequence he was forced to resign the chair of political economy at King's College, London, to which he had only quite recently been appointed.

For the next ten years Senior was constantly engaged in public affairs. He was chief adviser to the Whig party on economic problems, although

his advice was by no means always followed. A pamphlet he wrote in 1835 *On National Property and on the Prospects of the Present Administration and Their Successors*, can be regarded as an unofficial manifesto of the more advanced Whigs: it urged, *inter alia*, the reform of municipal corporations and the admission of dissenters to universities, and again recommended the transfer of some property of the Anglican church to the Roman Catholic church in Ireland. Senior was also a member of several commissions: the Poor Law Inquiry Commission, in 1833; the commission that investigated factory conditions, in 1837; the Royal Commission on the Distress of the Hand-loom Weavers, in 1841; and later, in 1857, the Royal Commission on Popular Education. Each time he was a member of a commission he wrote a substantial part of the commission's report. In his work for the Poor Law Commission, which won great praise, one of his closest collaborators was Edwin Chadwick.

With the fall of the Whig administration in 1841, Senior returned to more theoretical economics. In 1847 he was appointed to the Drummond chair of political economy at Oxford for a second five-year term. His intention of organizing his ideas into a treatise on political economy during these years was not to be fulfilled. Only a few of the lectures were published (1847; 1852), but a large part of the unpublished lectures of this period survive in manuscript. They are remarkable for their attack on the hypothetical economic man of Ricardo's imagination (Bowley 1937).

In the later years of his life Senior was a keen traveler, he was particularly interested in the revolutionary disturbances of 1848 in France and Italy, where he had numerous friends among the liberals. He also visited Greece, Turkey, Egypt, and Malta. He kept journals of these visits and of those to Ireland, some of which were subsequently published.

Senior's wide circle of friends included most of the leading Whigs of his day as well as such intellectuals and writers as Lord Jeffrey, Sidney Herbert, Macaulay, and Thackeray. Among the "philosophical radicals," his closest friends were the Austins and Grotes and, among men with whom he worked, Sir George Cornewall Lewis, Lord Howick, Chadwick, Sir James Stephen, and, of course, Archbishop Whateley. He knew well such French liberals as Alexis de Tocqueville, Guizot, and Faucher, and in Italy he was much appreciated by Cavour.

Senior was a Whig rather than a philosophical radical, but his conviction that man has a natural tendency to improve himself, provided unwise institutions do not discourage him, may well have

developed from his association with the disciples of Bentham. It was this view that led to his controversy with Malthus. Senior's proposals for poor-law reform included the abolition of outdoor relief to the able-bodied, for he believed this sapped the virtues of thrift and independence and also depressed wages. However, he hoped the destitute sick, the aged, and children would be looked after out of public funds in hospitals, almshouses, and orphanages, *not* in mixed workhouses. In general, he believed that the government had a duty to intervene whenever it could do good: expediency should determine the occasions, not a priori rules.

### Contribution to economic theory

During his life Senior's activity in the sphere of social and economic reform was more visible than his work on economic theory. His contribution to the latter, however, was considerable. In contrast to his younger and more famous contemporary, J. S. Mill, Senior's approach to Ricardian theory was essentially critical. His analysis of definitions in the appendix to Whateley's *Logic* (1826) shows how important precision in definition was to him. Concern with rigorous definition led him to make significant changes in the theory of value. Jevons was greatly impressed with this work and considered that he had anticipated the theory of marginal utility in important respects. Senior was clearly influenced by J. B. Say, and in connection with the analysis of degrees of competition in relation to value, he was also influenced by Samuel Bailey.

Senior was less successful with the theory of wages than with the theory of value. He realized that the wages-fund theorem was incomplete, since the determinants of the size of the fund had not been explained, but despite much effort he was unable to solve this difficulty.

His best-known contribution to economic analysis is his abstinence theory of capital. It emerged from his attempt to find a definition of capital that would make capital coordinate with labor in the explanation of the influence of real costs on value; this was an important step in the process of separating the labor theory of value from orthodox economics. The theory stated in his *Outline of the Science of Political Economy* (1836) became an integral part of English capital theory. It was accepted by Marshall as well as by Jevons; the former, however, substituted the neutral term "waiting" for "abstinence." The theory also influenced Continental and American economics.

In the *Outline*, Senior fully accepted Ricardo's theory of land rent while observing that the laws of increasing and diminishing returns are not logically symmetrical. He generalized the analysis to

cover all cases of factors in fixed supply, including natural abilities, and hinted at the concept of quasi rent. This resulted in difficulties in the identification of factor incomes and the relation of rents to costs. Influenced by the German economist Hermann's review of the *Outline*, Senior attacked the problem again in the 1847-1852 lectures. He accepted Hermann's view that land and capital are species of the same genus, arguing that they are indistinguishable in practice and that the distinguishing attribute of land is not its scarcity but its immobility (Bowley 1937).

These theoretical developments necessarily affected the treatment of capital. In the later lectures capital was defined as "any wealth destined to be used productively," including land. This productive use, i.e., capital accumulation, requires "abstinence" as well as the gifts of nature and industry. Abstinence was explained as "the conduct of a person who either abstains from the unproductive use of what he can command or designedly prefers the production of remote to that of immediate results." Frugality, Senior explained, is the "mere refraining from the use of a commodity," but "providence" is "the employment of labour to produce remote results" (quoted in Bowley 1937, pp. 160-161). It must be observed that Senior's distinction between frugality and providence, although apparently similar to the Keynesian distinction between saving and investment, was not made for the purpose of analyzing the level of economic activity.

Senior's contributions to the theory of money and international trade are of less interest today than his other analytical work, but they provided a number of useful clarifications and qualifications to the theory of his time. The discussions of scope and method were contained in various published lectures given at the time of his first and second Drummond professorships and in his *Outline*. His views developed considerably during his life, and the *Four Introductory Lectures on Political Economy* (1852) are a valuable contribution to the nineteenth-century literature on the subject.

MARIAN BOWLEY

[For the historical context of Senior's work, see the biographies of MALTHUS; RICARDO; SAY; for discussion of the subsequent development of his ideas, see the biographies of JEVONS; MARSHALL.]

### WORKS BY SENIOR

- 1821 Essay on the Application of Capital to Land, With Observations Showing the Impolicy of Any Great Restriction of the Importation of Corn, and That the Bounty of 1680 Did Not Lower the Price of It, by a Fellow of University College, Oxford. *Quarterly Review* 25:467-504. → Published anonymously.



- (1826) 1951 *Appendix: On Certain Terms Which Are Peculiarly Liable to Be Used Ambiguously in Political Economy*. Pages 227-239 in Nassau W. Senior, *An Outline of the Science of Political Economy*. New York: Kelley. → First published as an Appendix to Richard Whately's *Elements of Logic*.
- (1827) 1831 *Introductory Lecture on Political Economy*. 3d ed. London: Murray.
- (1828) 1931 *Three Lectures on the Transmission of the Precious Metals From Country to Country and the Mercantile Theory of Wealth*. Series of Reprints of Scarce Tracts in Economic and Political Science, No. 3. London School of Economics and Political Science.
- 1829 *Two Lectures on Population*. With a correspondence between the author and the Rev. T. R. Malthus. London: Saunders.
- (1830a) 1931 *Three Lectures on the Cost of Obtaining Money, and on Some Effects of Private and Government Paper Money*. Series of Reprints of Scarce Tracts in Economic and Political Science, No. 5. London School of Economics and Political Science.
- (1830b) 1959 *Three Lectures on the Rate of Wages*. With a preface on the causes and remedies of the present disturbances. New York: Kelley.
- 1831 *A Letter to Lord Howick: On a Legal Provision for the Irish Poor, Commutation of Tithes, and a Provision for the Irish Roman Catholic Clergy*. London: Murray.
- 1835 *On National Property and on the Prospects of the Present Administration and Their Successors*. London: Fellowes.
- (1836) 1951 *An Outline of the Science of Political Economy*. New York: Kelley.
- (1837) 1844 *Letters on the Factory Act as It Affects the Cotton Manufacture, Addressed to the Right Honourable President of the Board of Trade*. London: Fellowes.
- 1847 *A Lecture on the Production of Wealth*. Oxford: Combe.
- 1852 *Four Introductory Lectures on Political Economy*. London: Longmans.
- 1859 *A Journal Kept in Turkey and Greece in the Autumn of 1857, and the Beginning of 1858*. London: Longmans.
- Conversations and Journals in Egypt and Malta*. London: Low, 1882.
- Historical and Philosophical Essays*. 2 vols. London: Longmans, 1865.
- Industrial Efficiency and Social Economy*. Essays edited by S. Leon Levy. New York: Holt, 1928.
- Journals, Conversations and Essays Relating to Ireland*. London: Longmans, 1868.
- Journals Kept in France and Italy From 1848 to 1852: With a Sketch of the Revolution of 1848*. 2 vols. London: King, 1871.

## SUPPLEMENTARY BIBLIOGRAPHY

- BOWLEY, MARIAN 1937 *Nassau Senior and Classical Economics*. London: Allen & Unwin. → Contains a bibliography of Senior's published and unpublished writings.
- LEVY, S. LEON 1943 *Nassau W. Senior: The Prophet of Modern Capitalism*. Boston: Humphries.
- ROBBINS, LIONEL 1952 *The Theory of Economic Policy in English Classical Political Economy*. London: Macmillan.

- SCHUMPETER, JOSEPH A. (1954) 1960 *History of Economic Analysis*. Edited by E. B. Schumpeter. New York: Oxford Univ. Press.
- TAUSSIG, FRANK W. (1896) 1932 *Wages and Capital: An Examination of the Wages Fund Doctrine*. London School of Economics and Political Science.

## SENSES

- I. OVERVIEW  
II. CENTRAL MECHANISMS

F. A. Mote  
Ragnar Grant

I  
OVERVIEW

In any consideration of behavior it is important to have knowledge of the senses, for the first step in an organism's adjustment to its environment is sensory stimulation. The behavior of primitive organisms is simple and limited; this is so partly because their receiving mechanisms are of such a nature and so structured that the organisms can sense only gross changes in their restricted environment. Higher forms of life have more numerous, more complex, and more sensitive mechanisms; therefore, it is possible for them to react to more environmental changes as well as to changes of wider range and greater subtlety. At the human level the mechanisms have become so specialized, and the environment is so complex, that to understand much of man's most significant behavior, it is necessary to know about the receptor systems by which he senses and adjusts to his varied and changing world.

## Classification of senses

Although it is popularly believed there are only five senses, there are actually more. No specific number can be stated because it will vary from one listing to another, depending upon the criteria used to classify the senses. A set of criteria often used consists of the kind of subjective experience involved—visual, olfactory, tactile, etc.—the type of stimulus, and the sense organ, or receptor. Table 1 outlines this scheme.

## The nature of sensory experiences

The stimulation of a sense organ results in a unique kind of experience, and the differences in kind are what we identify in speaking of auditory, taste, and other types of sensations. The traditional five senses of sight, hearing, taste, smell, and touch have been defined this way since the time of Aristotle, but it has long been known that these do not make the complete list, since sensations of warmth, cold, heat, and pain—in addition to

**Table 1 — Scheme for classifying senses**

Subjective experience	Stimulus	Receptor
Vision	Electromagnetic radiations	Rods and cones of the retina
Hearing	Pressure of sound waves	Hair cells of the organ of Corti of the inner ear
Taste	Various substances in solution	Taste cells in taste buds of the tongue
Smell	Various vaporized materials	Bipolar neurons of olfactory nerve in olfactory epithelium in upper nasal cavity
Touch	Deformation of skin, mucous membrane, internal organs, etc., by stretching, pulling, and similar stimulation	Free nerve endings
Pressure ("deep," i.e., subcutaneous and internal pressure)	Pressure in excess of that required for touch, distension of internal organs, etc.	Pacinian corpuscles in subcutaneous tissue and internal organs
Warmth	Contact by stimulus above "physiological zero," radiant heat, elevation of temperature of surrounding medium	Receptor uncertain or disputed
Cold	Contact by stimulus below "physiological zero," lowering of temperature of surrounding medium	Receptor uncertain or disputed
Pain	Various electrical, thermal, mechanical, chemical, and other agents, e.g., electric shock, radiant heat, puncture, some acids, etc.	Free nerve endings
Vestibular (equilibrium, labyrinthine)	Linear and rotational acceleration, head and body position and movement	Cristae of semicircular canals and otoliths of macula of utricle and sacule (all parts of nonauditory portion of inner ear)
Kinesthesia (movement and "muscle sense")	Movement of muscles, tendons, and joints	Various receptors in muscles, tendons, and joints, e.g., spray-type nerve endings and free nerve endings in muscular tissue and joints, Golgi tendon organs, modified Pacinian corpuscles

touch—can be aroused when the skin is stimulated. Even when these are included, the list is still far too short because movement, pressure, fatigue, nausea, and many others that could be named are experienced as sensations just as much as, let us say, taste and sight. In addition to the differences in kind of sensation from one sense to another, there are also differences in quality within each sense. Not only is taste as a class of sensations different from sight, but the taste quality of sweetness is different from that of sourness, and the visual quality of whiteness different from that of other colors. The number of sensations and their variations and combinations is infinite. To order the domain of the sensations, Hermann von Helmholtz introduced the concept of modality, which is defined as the class of sensations, all of which are alike.

**Intensity.** The basic aspects of sensation are intensity, sensory modality, and sensory quality. First, let us briefly consider sensation intensity. Most sense organs are specialized; that is, they are composed of receptor elements whose threshold of excitability is lower for one form of energy than for any other. This particular type of energy is known as the adequate stimulus. Receptors will react to other stimuli—for example, the application of an electric current or pressure on the eyeball will arouse visual sensations—but the energy required is greater; such stimuli are often referred to as inadequate.

**Absolute threshold.** The minimum amount of the adequate stimulus necessary to excite a sense organ is called the absolute threshold. For some senses the energy required is extremely small; for instance, at the absolute threshold for human vi-



sion the rods can respond when they absorb only 6–10 light quanta, and it has been calculated that movements of the eardrum of the order of  $10^{-2}$  cm. can excite the auditory system.

**Differential threshold.** The senses, however, do more than detect the presence or absence of stimulation; they respond to differences in stimulus magnitude. As the value of the stimulus is increased above the absolute threshold, the sensation experienced becomes more intense—the light becomes brighter and the sound louder. But it is also important to know how much difference there must be so that one light can be seen as brighter than another or what change in sound frequency is required in order to be able to recognize that two tones differ. This is the problem of the differential threshold, defined as the just noticeable difference in stimulation. The size of the differential threshold is a measure of receptor sensitivity; when the just noticeable difference is small, sensitivity is high, when this difference is large, sensitivity is low.

**Work of Weber and Fechner.** E. H. Weber's research in the 1830s and 1840s led him to conclude that the just noticeable difference in stimulus intensities,  $\Delta I$ , depends upon the ratio of their magnitudes, that is,  $\Delta I = I_1/I_2$ , and that for a given class of stimuli the ratio is constant. Thus, if a light intensity of 11 units can be discriminated as just noticeably brighter than one of 10 units, then a light intensity must be of 110 units to be discriminated from one of 100 units, and so on—the constant ratio being  $1/10$  in this example. Although he knew the value of the ratio varied from one class of stimuli to another, Weber believed it was constant for any given class, a generalization that has usually been found to hold true for only a part of the total range of magnitude for any stimulus class. Gustav Fechner applied Weber's findings about intensities of stimuli to intensities of sensation, assuming that a just noticeable increase in stimulus intensity meant there was a constant increment in sensation intensity. He believed the quantitative relationship between stimulus and sensation was logarithmic, and in 1860 he expressed this in his famous psychophysical law:  $S = k \log I$ , where  $S$  is sensation intensity and  $I$  is stimulus intensity (Boring 1942). [A more detailed treatment of thresholds, the relationship between stimulus magnitude and sensation intensity, and the problems and methods involved in their measurement will be found in *PSYCHOPHYSICS*; see also the biographies of *FECHNER* and *WEBER*, *ERNST HEINRICH*.]

**Sensory modalities.** Although the importance of sensation intensity has always been recognized, most of the interest in the senses has centered

about the problems of sensory modality and the quality differences within a modality. Our daily experiences seem to provide the answer to the question "What is there about the stimulus, the sense organ, or the nervous system that explains sensory modality?" To common sense it appears self-evident that, although the nervous system transmits neural activity to the brain, where sensory experience occurs, the essential factor that accounts for sensory modality is the stimulus-sense organ relationship. This determines which experience takes place in the brain. Thus, the color of the grass stimulates the eye, and we have the sensation of greenness; the singing of the bird falls on the ear, and we hear; the skin senses the warmth of the sunshine and the cool of the shade. In order to experience a particular sensation, it is simply a matter of the natural stimulus exciting the appropriate sense organ.

However, it is not only the "natural" stimulus that can arouse a given sensation. For example, a blow on the head will evoke visual sensations and make the ears ring; the usual stimulus for dizziness is spinning or rotating the body, but rotating the visual field will have the same effect; a very warm stimulus may be felt as cold when applied to a cold-sensitive spot on the skin—the "paradoxical cold" phenomenon. These examples demonstrate an important fact: no matter how it is excited, whenever a specific sense organ is stimulated, the sensation is always the same. Yet it is not necessary for the sense organ itself to be stimulated in order for its appropriate modality of sensation to be experienced. For instance, when the optic nerve is cut, a flash of light is seen, and after amputation of an arm there may be reports of touch, itching, pain, and other sensations that are felt as coming from the missing hand and fingers. [See *BODY IMAGE*.]

**Doctrine of specific nerve energies.** If neither a particular stimulus nor stimulation of a particular sense organ is essential, then there must be something about the nervous system that is the necessary factor determining sensory modality. It is this conclusion that Johannes Müller stated in his doctrine, or law, of specific nerve energies (Boring 1942). It was Müller's hypothesis that each sensory nerve has its unique "energy," or quality. Müller's hypothesis declares that sound is sensed because of the unique specific energy of the auditory nerve, smell because of the specific energy of the olfactory nerve, and the other sense modalities because of similar specific energies—and the modalities differ because the specific nerve energies are different. Although no property of sensory nerves has ever been found which supports the specific-energy

concept, Müller's doctrine has been very influential in theory and research upon the senses. It led investigators to look for specialized receptors which, when they activate their sensory nerve, arouse a particular type of sensation [see MÜLLER, JOHANNES].

*Problem of special receptors.* For most of the senses an interrelationship can be found which involves an adequate stimulus, a specialized receptor, and a specific modality of sensation. Light stimulates the rods and cones of the retina, and we see; substances stimulate the taste buds of the tongue, and we taste. Relations of this sort can be found for many modalities. However, such information about some senses is uncertain, and our knowledge of how sensations evoked by stimulation of the skin are mediated is an example of this. The skin is not uniformly sensitive but has "spots" of sensitivity that respond to touch, temperature, etc. Under most conditions no temperature is felt at a given cutaneous region; it is at "physiological zero." Upon being stimulated point by point by a stimulus above this zero value, warmth is not felt at all points nor is cold when a stimulus below this value is used; warm and cold sensations are only reported upon stimulation at certain spots, and the number and location of these are different for the two kinds of stimuli. This finding has been interpreted as indicating the existence of a special receptor for warmth and a different one for cold at each temperature-sensitive spot. A similar interpretation has been proposed for touch, one type of receptor to serve for hairy skin and another type for hairless. Because of their number and nearness to pain-sensitive spots, the free nerve endings are considered to mediate pain. Although the free nerve endings usually are not considered to be specialized receptors, when they are appropriately stimulated, touch and temperature will also be sensed (Weddell 1955).

The cutaneous senses have sometimes been discussed in terms of a conception of modality which implicitly assumes there must be four morphologically distinct types of receptors. But the fact that all the cutaneous modalities can be evoked by stimulating the free nerve endings casts doubt upon the necessity to assume distinct receptor types. The meaning of cutaneous sensation also requires clarification. "Touch," "warmth," "cold," and "pain" do not designate unique sensations perceived in isolation; such terms refer to the predominant experience aroused by particular stimuli. A stimulus above physiological zero applied to the skin evokes warmth as the primary feeling, but the stimulus is also felt as a touch at a particular location, and it is perceived as having temporal and spatial charac-

teristics. This example also illustrates an important phenomenon, that of sensory interaction. The inputs from sensory systems interact at the higher nervous centers and are integrated there so as to give us our perceptions of the environment. [See SKIN SENSES AND KINESTHESIS.]

*Sensation quality.* To speak of sensation is also to speak of sensation quality; the two are inseparable. Visual sensations are white, or gray, or red, or yellow; a pain is "dull," or "aching," or "throbbing"; and similarly for all the senses. The usual method of investigating sensory quality has been to search for the aspects of the stimulus which correlate with quality and variations in quality. For many of the senses some of the most important correlations are known and the factors involved have been studied for many years. For vision the wave length of the light and color quality are correlated; sound frequency correlates with pitch, and the pattern of overtones with the characteristic timbre of musical instruments and the voice; lightly stroking the skin evokes the touch quality of tickling, and so on. Much of our knowledge about the senses consists of information of this sort, and many of the most important generalizations and theories about the senses, such as the Young-Helmholtz theory of color vision and the place theory of pitch for hearing, are explanations of sensory quality [see HELMHOLTZ].

*Receptor-quality correlations.* Obviously, before there can be any satisfactory scientific understanding, the receptor and the quality that are involved in the stimulus-quality correlation must be known. For some senses our knowledge about one or the other of these is questionable. In the case of thermal sensitivity the stimulation is temperature above or below physiological zero, and the sensory quality is warmth or cold, but since there is uncertainty about how thermal sensitivity is mediated, little that is conclusive can be said about any correlation. In the case of olfaction there is the opposite difficulty; here there is no question about the receptor, but anywhere from three to nine classes of odor qualities have been proposed. Since agreement has not been reached about the analysis of the qualities, here too no satisfactory stimulus-quality correlation can be established.

*Localization in the brain.* When it became clear that Müller's specific nerve-energy concept was not tenable, the only reasonable alternative hypothesis seemed to be one which he had considered as a possibility, namely, that when a sense organ is stimulated, its sensory nerve sets up excitation in a particular region in the brain; there is specificity because nerves from different sense organs go to different regions. Specificity does exist in the sense



that the optic-nerve radiations terminate in the occipital lobe of the brain, and hence this is the visual center; the auditory nerve ends in the temporal region, and this is the auditory center, and so on. However, although occasionally someone has entertained the possibility of something like them, no "green center," "warm center," or "center" for sweet tastes or flowery odors and the like has been found. Such a conception of specificity is too simple. [See NERVOUS SYSTEM, *article on* STRUCTURE AND FUNCTION OF THE BRAIN.]

### Methods and findings

Within this century the most significant contributions to the understanding of the senses have come from electrophysiological investigations. By appropriate techniques the responses of individual receptors and of single nerve fibers can be studied as the stimuli are varied. Under most conditions of stimulation the electrical changes recorded from a single nerve unit are a series of discharges, all of the same size, and the unit is therefore said to show all-or-none response. As intensity of stimulation is increased or decreased, it is not the magnitude of discharge of the nerve that varies, but its frequency; and when records are taken from several units, it is found that variation in stimulus intensity is also reflected in the number of units responding. The frequency of unit discharge and the number of units excited are considered to be the basic neural correlates of the dimension of sensation intensity.

Recent research on the electrophysiology of taste has provided data and information which show that some of the ways of looking upon sensation and sensation quality must be reconsidered. That the taste cell of the taste bud is the receptor for taste has long been known, and although from time to time different classifications of primary taste qualities have been proposed, it was finally agreed that there are four: sour, salty, sweet, and bitter. Most of the taste buds are found in small elevations on the tongue known as papillae. Individual papillae have been found which respond exclusively to salt, to sour, and to sweet, but most are stimulated by substances which evoke more than one quality. Recordings have been made of the electrical activity of single units of the nerves supplying various animal taste buds when these are being stimulated by weak acids, sodium chloride, sugar, quinine, and other substances that evoke the different primary taste qualities in the human subject. Such records show that most units respond to more than one substance, and from one unit to another there is variation in the degree of response to different substances. For example, one

unit might respond to both hydrochloric acid and sodium chloride, but in a different degree to each; another unit might respond to the same concentration of these, but to another degree for each, and in addition it might respond to quinine. A third unit might, again, respond to both the acid and the salt at the same concentration, and to still a different degree for each, and to sucrose as well. Thus, although many units may respond to stimuli for the primary taste qualities, each may have its own response "spectrum," or characteristic pattern of sensitivity. Such findings do not support a simple view of four basic receptor-cell or nerve-unit types corresponding to the four taste qualities. "Sweet," "sour," etc. are probably best considered as descriptive headings for taste qualities, rather than as terms signifying any sort of entity. It appears then that neither the responses of individual cells nor of single neural units determine taste quality—it is the patterning of sensory nerve activity which provides the basis for taste discrimination (Pfaffmann 1962).

Electrophysiological methods have been used to investigate the individual receptor elements, the nerve pathways, and the cortical centers for all the senses; information that could have been obtained in no other way has been gathered. By such methods new discoveries have been made which show that some of the traditional views and theories about the senses are too simple and inadequate. [See NERVOUS SYSTEM, *article on* ELECTROENCEPHALOGRAPHY.]

In the past the senses were studied for the purpose of explaining sensations and their qualities. Now there is less emphasis on this, and the senses are being studied in and for themselves in order to gain knowledge about their properties and their processes. Work undertaken with this new purpose is not only giving us better knowledge of sensations and qualities but it is also giving us a larger view of the senses and the part they play in the behavior of organisms.

F. A. MOTE

[Directly related are the entries HEARING; PAIN; SKIN SENSES AND KINESTHESIS, TASTE AND SMELL; VISION. Other relevant material may be found in ATTENTION; NERVOUS SYSTEM, *article on* STRUCTURE AND FUNCTION OF THE BRAIN; PSYCHOPHYSICS; and in the biographies of FECHNER; HELMHOLTZ; HERING; MÜLLER, JOHANNES; WEBER, ERNST HEINRICH; WUNDT.]

### BIBLIOGRAPHY

BORING, EDWIN G. 1942 *Sensation and Perception in the History of Experimental Psychology*. New York: Appleton.

- GELDARD, FRANK A. 1953 *The Human Senses*. New York: Wiley.
- PFAFFMAN, CARL 1962 *Sensory Processes and Their Relation to Behavior: Studies on the Sense of Taste as a Model S-R System*. Volume 4, pages 380-416 in Sigmund Koch (editor), *Psychology: A Study of A Science*. New York: McGraw-Hill.
- PIÉRON, HENRI (1945) 1952 *The Sensations: Their Functions, Processes, and Mechanisms*. New Haven: Yale Univ. Press; London: Müller. → First published as *Aux sources de la connaissance: La sensation, guide de vie*.
- SYMPOSIUM ON PRINCIPLES OF SENSORY COMMUNICATION, ENDICOTT HOUSE, 1959 1961 *Sensory Communication: Contributions*. Cambridge, Mass.: M.I.T. Press.
- WEDDELL, G. 1955 Somesthesia and the Chemical Senses. *Annual Review of Psychology* 6:119-136.

## II CENTRAL MECHANISMS

Physiology had very little precise information concerning central mechanisms of sensation until the advent of the micromethods of modern neurophysiology. The well-known basic facts concerning specific cortical projection areas for vision, hearing, etc., more properly belong to the conceptual sphere of functional anatomy and pathology. This article will attempt to make some sense out of the less accessible results dealing with the organization of sensory messages and their transmission as revealed by microtechniques in terms of single fibers and individual cells.

The central mechanisms of sensation come into operation at the first synapse which the neural message encounters on its way from the sense organ. In the eye and the ear, for instance, this synapse is located in the periphery at the base of the receptors themselves. For afferents leading from receptors in the skin, the synapse is found in the spinal cord and in its upper extension, the medulla oblongata. In the eye the message is twice reorganized before it reaches the optic nerve, at the synapse of the receptor and bipolar cells and at the synapse of the bipolar and ganglion cells. There, some "sharpening" of the information takes place, aided as we shall see, by the mobility of the eye. Skin and muscular afferents are synaptically connected in the spinal cord to elaborate reflex systems that in turn are controlled by internuncial cells (interneurons) run by centripetal circuits that can redirect, inhibit, or facilitate the passage of information.

The sense organs whose central mechanisms have been most carefully studied are the eye, the ear, and the muscle spindles. These will be used to illustrate some of the leading principles that have emerged from such work.

Let us first consider the extremely high sensitivity of the sense organs. If this sensitivity were

always fully utilized and its effect transmitted to the cortex, the higher centers would be at the mercy of an impulse barrage that would depolarize their neurons to a level fatal for organized life. Obviously situations exist in which full sensitivity is needed, as in extreme twilight vision for the eye, or in the minute adjustments of muscle length made by the spindle organs. But the eye must also be able to function in brightest sunlight, and most muscular movements are coarse and would be disturbed by excessive spindle discharges, which as we know, are returned reflexively to the muscles in which these organs are situated. These two examples suffice to show that sensitivity must be regulated. Much of the most important work on central afferent mechanisms has been devoted to the elucidation of both local and distant (or centripetal) controlling mechanisms that act *internally* on the neurons or receptors themselves. These are distinct from well-known *external* mechanisms, such as pupillary reflexes and activities of the stapedius and tensor tympani muscles in the middle ear.

An important restrictive *peripheral* mechanism requiring brief mention is the adaptation of the receptors that is expressed in a reduction of impulse frequency as a function of the length of time of stimulation (Adrian 1928). The so-called light and dark adaptation of the retina is a borderline case: there is a peripheral shift from rod, or scotopic, vision based on rhodopsin to cone, or photopic, vision as the level of illumination increases, but neural mechanisms within the retinal center itself are also deeply engaged in the adaptive adjustments of the sensitivity of the eye. The details of their mode of operation are not as well known as one could desire. Their existence, however, is well established by several workers (summarized in Granit 1962).

**Relays and transmission.** At an earlier time, the different synaptic stations traversed by a message were regarded merely as relays, a loosely defined function quite apart from the complications alluded to above. We realize today that afferent messages from different receptors at a relaying neuron may be mixed in different ways by the manner in which their afferent terminals converge and that controlling mechanisms interfere at each station, either to specify, amplify, or partially or wholly block a message. Nevertheless, there is always some relay function involved, in the sense that even a modified impulse message must be transmitted to the next station with some fidelity as to its "quantity." Since impulse frequency (in addition to the number of active cells) is a most important determinant of quantity or intensity, clearly there must be a mechanism for relaying it



with a good margin of safety. One factor making a relay reasonably reliable is "synaptic density." A relay could be made to transmit with an equal margin of safety without a dense synaptic projection, but in such systems facilitatory or inhibitory controls would be relatively more important in deciding under what circumstances maximum output would be permitted.

At each relay, impulse frequency is translated at the cell membrane into miniature synaptic currents which can be recorded (through microelectrodes introduced into the cell and connected to amplifiers) as miniature potentials (see Eccles 1957; 1964). These indicate depolarization when the input is excitatory and the opposite, or hyperpolarization, when the input is inhibitory. Several sense organs have both inhibitory and excitatory afferents, the classical example being the large spindle afferents with reciprocally connected inhibitory and excitatory terminals to the protagonist and antagonist ventral horn cells, respectively.

In an excitatory relay of substantial synaptic density, the miniature synaptic currents are in one direction, all depolarizing, and hence will produce a graded change, the depolarizing current, often recordable in terms of electrical potential. Eccles (1957; 1964) calls it the postsynaptic excitatory potential. It will be translated into an impulse discharge. Imitating this process by injecting current through the tip of an intracellular electrode, one finds impulse frequency directly proportional to current strength (Granit et al. 1963). At the next relay these impulses will produce an amount of excitatory current dependent upon the frequency of discharge and the synaptic density. Is this process also directly proportional to impulse frequency? Indirect evidence supports this proposition (see below). The relay function of an afferent neuron lying in the path of the message to the cortex thus consists in translating an impulse barrage into depolarizing current and that once again into an impulse barrage and so on alternately at each subsequent station. Since proportionality is maintained, the message can be transmitted without distortion of the relative order of magnitude. The lemniscal path seems to exhibit the simple relay function described (Mountcastle 1961). The proportionality constants relating depolarizing current strength to impulse frequency are known to vary, and thus if they happened to be large, a sensory effect could be stepped up by amplification, if not, the initial frequency of discharge may be reproduced at the end station in its original version.

When more complex functions are considered, the simple picture drawn above of a pure relay

function ceases to be sufficient. Transmission will then be greatly influenced by internuncial cells that "take orders" from elsewhere. A large number of them characteristically fire rapid bursts of impulses when stimulated by a single shock. When tested for their impulse-producing capacity by injected currents, these cells have been proved to possess very large proportionality constants, meaning that very small amounts of current elicit fast frequencies of discharge. Such cells are therefore well suited for "biasing," in the positive or negative direction, the relays upon which they impinge.

**Specific and diffuse afferents.** When a sense organ has to deliver specific information of a more discriminative nature referring to spatial localization or some modality, this information is arranged spatially with considerable precision both at intermediate and at end stations. Topographic differentiation by means of somatotopic afferents is thus a major analytical tool of the physiological organization by which input is differentiated. In vision the physiological maps are particularly precise. The eyes, for instance, are represented by alternate layers in the geniculate body in such a manner as to join corresponding points of the retinal field in identical layers. The information, thus organized, is delivered to a cortical map of great precision. For touch, Woolsey (1952) has worked out detailed cortical charts by the method of evoked potential. Even quality may in some systems be mapped out. This is the case in the cortical area for tonal representation (Tunturi 1950). Examples could be multiplied to show how information on the structure of bodily space is charted in this manner by localized signals.

Differentiation by the basic principle of maintaining the topography of sensory impulses from sense organ to cortex is, however, not the only afferent mechanism known. There are also cells which serve diffuse or nonspecific functions. This concept is not easy to define with precision. Perhaps the best way of defining a diffuse afferent system is to call it "a system of cells on each of which afferent fibers of several modalities converge," or if the afferents be of one modality, "cells receiving projections from an extensive section or area of the body." Magoun's (1950) concept of the ascending reticular system is of the former type, in which neurons in the brain stem receive terminals from afferents representing a large number of different modalities, as proved in unit analysis by Moruzzi and his co-workers (Scheibel et al. 1955). Of the other type is the spinothalamic system, whose thalamic neurons have receptive fields which may include one half or the whole of the

body (Mountcastle 1961). They may, however, have terminals from two modalities (if pain be a modality) because nociceptive stimuli also discharge the cells which are sensitive to mechanical stimulation of the skin. In the retina the giant ganglion cells have branches that collect information across an area with a diameter of maximally 1 mm., representing several hundred thousand receptors, both rods and cones. Within the tactile system of the gracilis nucleus of the dorsal-column afferents, there are separate cells for localized messages and for highly convergent information coming from a large area.

What function should be assigned to diffuse systems? Only a partial answer can be given. For the reticular activating system, Starzl, Taylor, and Magoun (1951) suggest (as does the name itself) that it plays the role of keeping the brain alert and active, since somnolence occurs when this portion of the brain stem is removed. Granit (1955), drawing attention to the widespread spontaneous activity of nerve cells, has made two suggestions: (1) that the diffuse sensory input serves an energizing function; that is, it maintains some basic depolarization by random impulses which is necessary for the upkeep of spontaneous activity in the higher centers; and (2) that spontaneous activity in its turn is necessary as a background for the central recognition of inhibition. As we shall see below, inhibition is necessary in discrimination, but it needs a background of central excitation in order to be "informative." In studying intracellularly the afferent input from stretch receptors, we have been able to demonstrate (Granit et al. 1964) that, unless the cell (a motoneuron) is depolarized to some extent, both inhibitory and excitatory effects of stretch tend to be dissipated as sheer noise consisting of miniature potentials at the cell membrane, while in slightly depolarized cells, excitations really excite and inhibitions really inhibit.

The final cell. The neurophysiologist whose understanding depends on interpreting "spikes" (impulses) or membrane events (synaptic activation noise, membrane potential) has no experimentally based philosophy as to which, if any, of his indices comes close to the conceptual world of psychology. So far he has done best on spikes and has been tacitly inclined to imagine that unless a cortical cell delivers spikes, it is of but modest interest in the nervous machinery designed for discrimination of one message from another. Yet, while he is recording from the "final cell" in a cortical layer, no neurophysiologist is likely to believe that he is doing anything more than sampling one link in an organization of cortical cells that, ac-

cording to the anatomist Ramón y Cajal (1923), consists of a "cortical sector" (a "cone" would have been a more appropriate term) with its top downwards. Within this sector the cells are linked vertically. Physiological research (see below) has recently been able to make valuable use of Ramón's concept.

The role of inhibition. Impulse recording made it possible for the first time to assign definite tasks to nervous inhibition in the elaboration of the sensory message. Previously, central inhibition had been measurable only in reflex work, using muscular contraction or secretion as indices. Graded excitation (depolarizing current) has its equivalent antagonist in graded inhibition (hyperpolarizing current), often recordable by the intracellular technique as the "inhibitory postsynaptic potential" (Eccles 1957; 1964). However, most information of interest for the present theme has been obtained in terms of a reduction of impulse frequency.

Inhibition within a sensory system achieved functional prominence in the vertebrate retina for the first time when it was shown that the off-discharge at cessation of illumination was preceded by inhibition and easily inhibited by reillumination (Granit & Therman 1935; Hartline 1938). The retina is our most highly developed sense organ, and if natural stimulation of it by light could stop a discharge as well as start it, then surely inhibition is of first importance. The on-off pattern of discharge delivered by the individual optic nerve fibers was later also seen in the auditory system (Bremer 1943). However, in vision its significance is far better understood, apparently because the spatial element of the organ of sight is more accessible to analysis. From the beginning it was held that the on-off units of the retina would blaze the trail of a moving object with a recognizable spike pattern and that, similarly, contours would be emphasized because both regular and small, irregular eye movements light it up with on-off spikes. It has since been shown in precise experiments (Ditchburn 1955) that if the image on the human retina is maintained stationary by appropriate optical arrangements compensating for slight, unavoidable eye movements, then it also tends to fade out.

The on-off element, and likewise a smaller number of pure "on" and pure "off" elements, represents a receptive field of convergence together with an analyzer residing in the nervous structure between receptors and the ganglion cell whose message is recorded. The size and organization of the receptive field were first studied, with the aid of small spots of light, by Hartline (1940) in the frog and



then in the cat (Kuffler 1953), where they were found to be differentiated to the extent that they possess a center that is either "on" or "off." In both cases the surrounding part of the field had the opposite character. In the goldfish the receptive field has also been investigated with respect to wave-length sensitivity, and it has been found that this antagonism between center and periphery with respect to on and off may also be an antagonism in terms of spectral regions of sensitivity. The overlapping mosaic of receptive fields thus forms a dynamic interaction pattern which, transmitted up the optic nerve, can do justice to the infinite variety of form, movement, and color that a moving animal has to interpret and that it ultimately reproduces in the form of a stable world of sight. In this act inhibition is just as vital as excitation.

**Recurrent inhibition.** Recurrent inhibition is another significant and apparently common process in the organization of the sensory (and motor) message because it rejects unwanted components, emphasizes differences of "quantity" within a complex message, and apparently also stabilizes a steady discharge. Its basis is the common occurrence in all centers of recurrent collateral fibers returning directly or across an interneuron to the cell or cell system from which the parent fiber emerged. Most of our knowledge of this process, inasmuch as it consists of precise measurements, derives from cells as different as vertebrate ventral horn cells and the horseshoe crab (*Limulus*) ommatidia. In the latter case (Hartline & Ratliff 1958) the afferent large axon from each ommatidium sends fibers with inhibitory axo-axonic synapses to adjacent axons. Assume that an image is focused onto some ommatidia with irradiated light unavoidably spreading across surrounding ones. The ommatidia in focus will fire at higher frequencies than those in the surrounds. The recurrent inhibition has proved to be proportional to impulse frequency and hence the focal cells will inhibit those in the neighborhood more than the latter can inhibit them. As a consequence the image will be sharpened by "contrast."

Nature has preserved, as it were, the idea of recurrent, or "lateral," inhibition (in Hartline's terminology) through countless ages of phylogenetic development; recurrent fibers are found in most centers, both motor and sensory. In the ventral horn cells (Granit & Renkin 1961), in addition to "motor contrast," a stabilizing effect on the discharge frequency in tonically active cells can be demonstrated. While in the *Limulus* the mechanism appears to be wholly automatic, in the ventral horn cells it is provided with an internuncial neu-

ron, the so-called Renshaw cell, and hence is facultative (for a summary, see Granit 1963).

These cases of recurrent inhibition are the only ones analyzed in detail, but further developments can be expected, since qualitative work has already been done, for instance, in the hippocampus, in the olfactory bulb, and in the gracilis nucleus.

Lateral inhibitory effects delimiting the area stimulated have also been described under the heading "surround" or "afferent inhibition" by Mountcastle (1957) and his co-workers. Thus Powell and Mountcastle (1959), recording from the somatic cortical area of the macaque monkey, have found skin fields in which a central nucleus of excitation to touch was surrounded by an inhibitory zone. The mechanism in this case may well be recurrent inhibition improving at each relay definition of locality.

Inhibition also apparently plays a decisive role in sharpening up information about quality. In some optic nerve fibers the spectral response in terms of threshold sensitivity appears restricted to, say, the red, green, or blue wave lengths. This is the so-called modulator type of response. The dominator type of response is sensitive across the spectrum. The modulator bands are now generally held to be caused by inhibition sharpening up color specification. Broad and narrow bands of spectral sensitivity have also been recorded from higher visual stations. A recent discussion of the neural mechanisms specifying wave length has been given by Granit (1962). As to specification of tonal quality, Galambos (1954) points out that inhibition in this system also restricts the originally broad sensitivity of the peripheral acoustic mechanism to a narrow band, a process that is likely to take place already within the cochlea. Tonal bands of different width, also depending upon the frequency range which the nerve fibers carry, are thus delivered to both the visual and the auditory cortical area.

**Centrifugal control.** The sense organ in which centrifugal control is best understood is the muscle spindle; this organ is located within a spindle of thin muscles (so-called intrafusal muscle fibers) that are provided with special efferent fibers that can be stimulated both artificially and across natural reflex or supraspinal connections. Under such circumstances the intrafusal muscle fibers contract and thereby stretch the sensory spindle endings, which respond by discharging afferent impulses. There is a large literature in this field (Leksell 1945; Granit 1955) which cannot be reviewed here. Relevant in this connection is the principle of centripetal determination of the level of sen-

sitivity in terms of spindle length. This is of fundamental importance because from the spindle the afferent impulses run to ventral horn cells, setting up efferent impulses to—among others—the large muscles in which the spindles are located; these muscles in turn are forced to contract until the tension on the sense organ is relieved. Thus, the length at which the muscle operates is determined by centrifugal control of spindle length, an interesting case of a centrally controlled sense organ which at the same time is a motor organ.

There is also a considerable literature on the olivo-cochlear Rasmussen bundle of efferent fibers, which exert a direct inhibitory effect that can be recorded both in auditory single fibers and at the cochlea. Both crossed and uncrossed olivo-cochlear fibers exert inhibitory effects, but their significance cannot yet be said to be fully understood. A possibility is that they play a role in binaural localization. (For the relevant literature see Galambos 1956; Fex 1963.)

The retina is likewise provided with centrifugal fibers, which is hardly surprising because it is a true nervous center in spite of its peripheral localization, and as pointed out above, centripetal control of nervous centers is common. The scant literature in this field has been summarized by Granit (1962).

**Cortical analysis.** At the level of detail alluded to in this brief review, analyses are restricted to the study of the organ of sight (see Hubel & Wiesel 1959; 1963; Jung & Kornhuber 1961). Hubel and Wiesel have made a thorough investigation of the receptive fields of cortical cells stimulating the eye (of a cat) by small visual objects influencing the firing of single cortical units. The receptive fields in the cortex were found to be organized either with excitatory on- or inhibitory off-centers in the middle, the surrounds being of opposite character (see above for similarly organized retinal receptive fields). The cortical receptive fields are more elongated than the retinal ones, and they are oriented in different directions. The antagonism between center and periphery of the cortical receptive field makes stimuli covering the whole field or diffuse illumination relatively ineffective. The same properties make these fields extremely sensitive to form, size, position, and orientation of the stimulus and likewise sensitive to directional movement of a spot of light across the retina; this is easily understood, considering that a moving spot can traverse elongated fields in different directions. Some units were binocular, others monocular.

Recalling that Powell and Mountcastle (1959) had found that tactile cortical regions are columnar in shape with similar responses to a microelectrode which apparently penetrates one of Ramón's sectors (see above) from above downward, Hubel and Wiesel (1963) made an analysis of receptive visual fields in a similar manner. They found a number of columns in the visual area in which receptive-field orientation was maintained, suggesting that columns of ordered sequences of field orientation are actually part of the central sensory mechanism, defined, of course, in functional terms.

RAGNAR GRANIT

[See also NERVOUS SYSTEM, articles on STRUCTURE AND FUNCTION OF THE BRAIN AND ELECTROENCEPHALOGRAPHY. Other relevant material may be found in HEARING; PAIN; PSYCHOLOGY, article on PHYSIOLOGICAL PSYCHOLOGY; SKIN SENSES AND KINESTHESIS; TASTE AND SMELL; VISION.]

#### BIBLIOGRAPHY

- ADRIAN, EDGAR D. (1928) 1934 *The Basis of Sensation: The Action of Sense Organs*. London: Christophers.
- BREMER, FRÉDÉRIC 1943 *Étude oscillographique des réponses sensorielles de l'aire acoustique corticale chez le chat*. *Archives internationales de physiologie* 53: 53-103.
- DITCHBURN, R. W. 1955 *Eye-movements in Relation to Retinal Action*. *Optica acta* (London) 1:171-176.
- ECCLES, JOHN C. 1957 *The Physiology of Nerve Cells*. Baltimore: Johns Hopkins Press.
- ECCLES, JOHN C. 1964 *The Physiology of Synapses*. Berlin: Springer.
- FEX, J. 1963 Crossed Cochlear Efferents Activated by Sound Through Both Ears. *Acta physiologica Scandinavica* (Stockholm) 59 (Supplement 213):41 only.
- GALAMBOS, ROBERT 1954 Neural Mechanisms of Audition. *Physiological Reviews* 34:497-528.
- GALAMBOS, ROBERT 1956 Suppression of Auditory Nerve Activity by Stimulation of Efferent Fibers to Cochlea. *Journal of Neurophysiology* 19:424-437.
- GRANIT, RAGNAR 1955 *Receptors and Sensory Perception*. New Haven: Yale Univ. Press.
- GRANIT, RAGNAR 1962 The Visual Pathway. Pages 535-796 in Hugh Davson (editor), *The Eye*. Volume 2: The Visual Process. New York: Academic Press.
- GRANIT, RAGNAR 1963 Recurrent Inhibition as a Mechanism of Control. Pages 23-37 in *Progress in Brain Research*. Volume 1: Brain Mechanisms. Amsterdam and New York: Elsevier.
- GRANIT, RAGNAR; KELLER, J. O.; and WILLIAMS, T. D. 1964 "Adjacent" and "Remote" Postsynaptic Inhibition in Motoneurons Stimulated by Muscle Stretch. *Journal of Physiology* 174:453-472.
- GRANIT, RAGNAR; KERNELL, D.; and SHORTESS, G. K. 1963 Quantitative Aspects of Repetitive Firing of Mammalian Motoneurons, Caused by Injected Currents. *Journal of Physiology* 168:911-931.
- GRANIT, RAGNAR; and RENEIN, BARBARA 1961 Net Depolarization and Discharge Rate of Motoneurons, as Measured by Recurrent Inhibition. *Journal of Physiology* 158:461-475.
- GRANIT, RAGNAR; and THERMAN, P. O. 1935 Excitation and Inhibition in the Retina and in the Optic Nerve. *Journal of Physiology* 83:359-381.



- HARTLINE, H. K. 1938 The Response of Single Optic Nerve Fibers of the Vertebrate Eye to Illumination of the Retina. *American Journal of Physiology* 121:400-415.
- HARTLINE, H. K. 1940 The Receptive Field of the Optic Nerve Fibers. *American Journal of Physiology* 130:690-699.
- HARTLINE, H. K.; and RATLIFF, FLOYD 1958 Spatial Summation of Inhibitory Influences in the Eye of *Limulus*, and the Mutual Interaction of Receptor Units. *Journal of General Physiology* 41:1049-1066.
- HUBEL, D. H.; and WIESEL, T. N. 1959 Receptive Fields of Single Neurons in the Cat's Striate Cortex. *Journal of Physiology* 148:574-591.
- HUBEL, D. H.; and WIESEL, T. N. 1963 Shape and Arrangement of Columns in Cat's Striate Cortex. *Journal of Physiology* 165:559-568.
- JUNG, RICHARD; and KORNHUBER, HANS (editors) 1961 *Neurophysiologie und Psychophysik des visuellen Systems. The Visual System: Neurophysiology and Psychophysics*. A symposium. Berlin: Springer. → Contributions in English, German, and French.
- KUFFLER, STEPHEN W. 1953 Discharge Patterns and Functional Organization of Mammalian Retina. *Journal of Neurophysiology* 16:37-68.
- LAFORTE, Y.; LUNDBERG, A.; and OSCARSSON, O. 1956 Functional Organization of the Dorsal Spino-cerebellar Tract in the Cat. II: Single Fibre Recording in Flechsig's Fasciculus on Electrical Stimulation of Various Peripheral Nerves. *Acta physiologica scandinavica* (Stockholm) 36:188-203.
- LEKSELL, L. 1945 Action Potential and Excitatory Effects of Small Ventral Root Fibres to Skeletal Muscle. *Acta physiologica scandinavica* (Stockholm) 10 (Supplement 31):1-84.
- MAGOUN, H. W. 1950 Caudal and Cephalic Influences of the Brain Stem Reticular Formation. *Physiological Reviews* 30:459-474.
- MOUNTCASTLE, VERNON B. 1957 Modality and Topographic Properties of Single Neurons of Cat's Somatic Sensory Cortex. *Journal of Neurophysiology* 20:408-434.
- MOUNTCASTLE, VERNON B. 1961 Some Functional Properties of the Somatic Afferent System. Pages 403-436 in Symposium on Principles of Sensory Communication, Endicott House, 1959, *Sensory Communication*. Cambridge, Mass.: M.I.T. Press.
- POWELL, T. P.; and MOUNTCASTLE, V. B. 1959 Some Aspects of the Functional Organization of the Cortex of the Postcentral Gyrus of the Monkey: A Correlation of Findings Obtained in a Single Unit Analysis With Cytoarchitecture. *Johns Hopkins Hospital Bulletin* 105:133-162.
- RAMÓN Y CAJAL, S. 1923 Studien über die Schrinde der Katze. *Journal für Psychologie und Neurologie* 29:161-181.
- SCHIEBEL, M. et al. 1955 Convergence and Interaction of Afferent Impulses on Single Units of Reticular Formation. *Journal of Neurophysiology* 18:309-331.
- STARZL, T. E.; TAYLOR, C. W.; and MAGOUN, H. W. 1951 Collateral Afferent Excitation of Reticular Formation of Brain Stem. *Journal of Neurophysiology* 14:479-496.
- TUNTURI, ARCHIE R. 1950 Physiological Determination of the Arrangement of the Afferent Connections to the Middle Ectosylvian Auditory Area in the Dog. *American Journal of Physiology* 162:489-502.
- WOOLSEY, CLINTON N. 1952 Patterns of Localization in Sensory and Motor Areas of the Cerebral Cortex. Pages 193-206 in Milbank Memorial Fund, *The Biology of Mental Health and Disease*. New York: Hoeber. → A discussion of Woolsey's article appears on pages 206-225.

## SENSORY AND MOTOR DEVELOPMENT

Sensory and motor functions are basic to all behavior. In its simplest form, the stimulus-response unit of behavior is composed of a receptor (sense organ), neural impulses traveling over afferent, central, and efferent pathways, and some resultant form of motor (muscular) response. The sense organs respond selectively to various stimuli (visual, auditory, olfactory, thermal, tactual, proprioceptive, chemical, and gravitational), and the resulting responses most often involve some form of muscular reaction appropriate to the nature of the stimuli received. Developmentally, the earliest forms of behavior are simple sensorimotor reflexes. With growth, maturation, and differentiation, the senses become more acute, and the muscles become stronger and function more smoothly. At the same time, the central nervous system matures, with its increasingly meaningful accompanying sensory perceptions, and the motor responses become better organized, while many of the reflexes give way to behaviors under voluntary control.

During the period of earliest development, when the changes are most rapid, there is a close coordination between structure and function. Gradually the functions become relatively independent of the structures. As the rate of growth slows, the structures differentiate, and the functional processes become increasingly complex; that is, once the basic structures are formed, there is little or no correlation between their normal variations in structural complexity or maturity and the increasing complexity and diversification of motor coordinations, perceptions, and other mental processes.

Neural and cortical development. Conel's studies of the postnatal development of the human cerebral cortex (1939-1963) have been well summarized by Eichorn (1963) and by Eichorn and Jones (1958), who also point out that changes in the histologic structure coincide with developmental changes in neural function as expressed in the electroencephalogram. At birth and even at one month the cortex is very immature, with fragile cell processes, no Nissl bodies, and very few neurofibrils. The greatest cortical development occurs

in the primary motor area of the upper trunk, leg, hand, and head, followed in order by primary somesthetic, visual, rhinencephalon (olfactory), and auditory areas, with other parts still very immature. By three months there are marked advances in the number of nerve fibers, both exogenous and associational, with greatest development in the motor area of the hand, frontal eye fields, and striate cortex (Gruner 1962). There is also over this period a rapid advance in myelination of the neural fibers. This myelination serves to channel the neural impulses along fibers and to reduce random spread of impulses across neurons. Again, at six months there is marked development, particularly in motor areas controlling the hand and upper trunk, leg, and head, while visual and somesthetic sensory growth is accelerated. Between 6 and 15 months the motor areas of the brain show less marked growth, with the order of maturity being hand, upper trunk, head, and leg. The primary visual area by now is second to the motor, with the visual association areas more developed than the somesthetic association areas.

Parallel with these histological changes, Eichorn and Jones (1958) point out that at birth the electrical activity of the cortex is very slow and irregular, with the greatest regularity in the region of the most mature cortical structure. It is possible, however, to induce some rhythmic EEG patterns in the neonate and in the month-old infant, while between one and three months there is a shift of the EEG from random activity to some patterned slow activity in the visual and auditory sensory areas of the brain.

**Fetal development.** We find, too, a definite parallel between early behavior and the neural histology and electrical functions in these early months. The very first actions of the fetus, according to Hooker (1943), are muscular: the rhythmic beating of the heart in the third week of gestation. This, however, is a preneural action of the heart muscle. A response (presumably neural) to stimulation was first observed at eight weeks and consisted of a lateral bending of the neck which moved the head away from a hair touching the area of the cheek. Carmichael (1946) has given an excellent account of this early fetal development. He points out the gradual involvement of the entire body and the appearance of reflexes until, by 26 weeks of gestation, the reflexes necessary to life are usually present. These reflexes include functioning of the respiratory, circulatory, and digestive systems as well as the sense organs that respond to light, sound, touch, body position, and so on.

### Sensory development

Because the infant's repertoire of responses is so limited, it is difficult to obtain exact information about sensory acuities. However, it is possible to observe and record such behaviors as visual regard, pupillary reflexes to light, startle, and changes in activity level to sounds and tactile stimulation. More recently, sensory reactivity has been recorded by observing changes in EEG and in heart rate and by such devices as observing eye nystagmus to moving striped patterns (Eichorn 1963; Fantz & Ordy 1959).

It is evident that the intact full-term newborn in some degree sees, hears, and responds to pressure, touch, taste, and change in temperature. There is evidence from his behavior and from the structures of the nervous system that of his various senses, vision is most developed.

**Vision.** Changes in visual acuity during the first month appear to be very slight. As observed in a standard test of infant development, soon after birth the infant will briefly regard a large moving object (such as a person) nearby and directly in his line of vision. A little less often he will regard a small bright red object in motion, when it is held about eight inches above his eyes (Bayley 1933; White, Castle, & Held 1964). At about two weeks his gaze may follow this moving object (a red plastic ring) across his visual field—right to left or the reverse (Bayley 1933). At three weeks his eyes may follow a moving person two or three feet away. At about one month he follows the red ring with up and down eye movements and, a little later, as it is moved slowly in a circle (18 to 24 inches in diameter). At six or seven weeks the infant appears to inspect his surroundings when carried in an upright position, and he turns his eyes toward the red ring at a thirty-degree angle when it is moved into his field of vision from the side. By the fourth month the infant's retina is able to accommodate to objects at varying distances in an almost adult fashion (Haynes, White, & Held 1965).

In experimental situations several investigators have found very early evidences of differentiation of visual stimuli. Several studies have shown (Berlyne 1958; Fantz 1958) that infants three to four months of age indicate preference for (that is, spend more time looking at) patterned stimuli as contrasted with plain ones. Fantz and Ordy (1959) have shown that infants under five days of age will look more at black and white patterns than at plain-colored surfaces. Doris and Cooper (1964,



p. 456) have reported a clear correlation between age and brightness discrimination among 16 infants 4 to 69 days of age. They tested this by observing nystagmic eye movements to a moving field of black and white stripes. [See PERCEPTION, article on PERCEPTUAL DEVELOPMENT; VISION, article on EYE MOVEMENTS.]

These findings are in general agreement with the responses to visual stimulation observed in the infant mental scales. Continuing with the Bayley Scale (Bayley 1933), at around two months the baby blinks at the shadow of a hand passed quickly across his eyes, he visually recognizes his mother, and his eyes follow a moving pencil; at two and one-half months he searches with his eyes for a sound, and he regards a one-inch red cube on a table when he is held upright; at three and one-half months his eyes follow small objects, such as the red ring, a teaspoon, and a ball, as they move across the table before which he is held in a sitting position. A typical four-month-old's occupation is to inspect his own hands; at four and one-half months he regards a pellet one-quarter inch in diameter; at five months he discriminates between strangers and familiar persons (largely visually, it would appear, from the expressive nature of his gaze). This evidence of visual discrimination of patterned objects shows advancement when at twelve months he looks with interest at colored pictures in a book. Many of his behaviors in the second year give evidence of his utilization of visual discriminations as he imitates motions, builds towers of cubes, adjusts round, square, and triangular blocks into their appropriate form-board holes, and goes on to more complex operations.

Increasing visual acuity in the first few months of life for premature and full-term infants has been assessed by Brown (1961).

Another source of information on visual development comes from the studies of ophthalmologists. Keeney (1951) has tabulated functional development of vision and binocularity for a series of ages from the third fetal month to nine years. Many of his items are identical with, or closely similar to, those already noted. We may add sensitivity to light at the seventh fetal month and a series of visual acuity fractions starting at one year, when visual acuity is about 6/60 with imperfect fusion. At two years it is at least 6/12. At two and one-half years more mature mechanisms of accommodation result in improved acuity. At three years vision is about 6/9, at three and one-half years fusion capacity is improving, at four years vision is near 6/6. At five years ocular pursuit is inferior to fixation, and at five and one-half

years fusion is well established and accurate. By six to six and one-half years ocular pursuit is accurate, and the average child can discriminate letters and word symbols and begin to read. Thereafter up to the age of nine, ability to tolerate prism vergences develops and continues to increase. [See VISION.]

Thus, we see that even though vision is present at birth and, relative to the other senses, advanced, acuity in one aspect or another continues to increase, at least up to nine years. The changes are more rapid at first and become slower with advancing age.

The eye is the most highly developed and complex of the sense organs, and we find, accordingly, that the development of visual acuity is a function of several variables, including the simple ones, brightness and hue; patterned vision, which is related to degree of complexity of both qualitative and quantitative variables; and depth discrimination, both monocular and binocular, together with the development of accommodation and convergence. Much remains to be done in clarifying and identifying the developmental aspects of all of these.

In the senses generally, and most acutely in vision, the pure sensory aspects of development are confounded with perception and the meaningful and adaptive responses to the stimuli which are presented for the study of sensory discrimination.

**Hearing.** The developmental pattern of auditory acuity is in many ways similar to that of vision. The normal newborn infant responds, by reflex startling, to a sharp, loud clack or the ringing of a bell near his ear (Bayley 1933). Ten days after birth he reacts to the lesser sound of a rattle and at twenty days to the sound of a softly speaking voice. The localization of the source of a sound is incipient in the two-and-a-half-month-old who may be seen to search with his eyes for the bell or rattle when sounded outside his field of vision. By three months his eyes will turn from the bell to the rattle and back, when they are sounded alternately while held eight inches apart in his field of vision. The four-month-old will turn to the right and the left to see the bell which has been rung opposite first one ear and then the other.

There is evidence of rapid cortical development in the auditory area of the cortex in the first four months. Normal acuity appears to be well developed by this age (Wever 1949). Further changes in hearing appear to be primarily perceptual. The six- to seven-month-old is interested in producing sound. He bangs his hand or a toy on the high-chair and listens. He babbles and repetitiously tries

out syllables. At eight to nine months he listens selectively to familiar words and begins to respond appropriately to simple commands.

Auditory acuity varies according to the pitch of the sound (Wever 1949, p. 364). However, this variation in pitch sensitivity appears to be a function of individual differences rather than development in the infant and child. After thirty years there is some decrement in auditory acuities, and this is greater for increasingly higher pitches (Wever 1949; Sommers, Meyer, & Fenton 1961). [See HEARING.]

**Tactual and pain sensitivity.** There is evidence of increasing sensitivity to pain in the first four days of life (Lipsitt & Levy 1959) and probably for a somewhat longer period of infancy. Schludermann and Zubek (1962) found no changes in pain sensitivity from age 12 through 50, though decrements occurred after that age.

The young infant clearly reacts to tactual stimulation. However, skin sensitivity also appears to increase with age. For example, Ghent (1961) studied tactual thresholds in the hands of children 5 to 11 years old and found that sensitivity increased over this age range. She also found a sex difference, with girls showing greater sensitivity and approaching the adult level of sensitivity at an earlier age. [See PAIN; SKIN SENSES AND KINESTHESIS.]

## Motor development

The development of motor coordinations, evidenced first in simple reflexes, appears to depend on the interactions of muscular response to stimulation, growth and increasing strength of the muscles, and the development of coordination through practice. All of these are interdependent. Practice strengthens the muscles and stimulates their growth. It also promotes learning, for example, through the simultaneous stimulation of visual and muscle senses in the eye-hand coordinations involved in reaching, grasping, and manipulating small objects. In newborn infants these coordinations are seen in such reflex responses as head lifting, various postural adjustments to body position, crawling, and reflex grasping. Soon, between one and two months, we observe playful bursts of activity in the form of arm and leg thrusts. As the muscles grow stronger, the infant is able to hold his head erect, to push his chest up by his arms, to turn from his back to his side at four months, and to sit, at first with support and by six months, alone (momentarily). By three months his hands are no longer tightly fistled, and he holds a small toy with a grasp which is no longer entirely reflex.

The six-month-old will reach for a toy with one hand. (Earlier he tends to "close in on" an object, using both hands simultaneously.) He shows early manual coordination in rotating his wrist, in partially using his thumb in opposition to his fingers, in grasping, and in trying to pick up pea-sized pellets. The eight-month-old sits alone steadily, may be starting to crawl or creep, and picks up small objects with complete thumb opposition. By nine months he can get himself into a sitting position and pull to a standing one by his crib rail. The ten-month-old creeps with agility and can often walk with help, sit down, and bring his hands together for games like pat-a-cake.

The one-year-old can take a few steps alone. In the next six months he will be able to throw a ball, walk backwards, and walk up and down stairs with help. The two-year-old walks up and down the stairs without holding on, and by three years he jumps from small heights, runs, walks on tiptoe. The four-year-old can walk a line and can hop a few steps on one foot.

**Individual variability.** There are of course, large individual differences in the age at which children become able to do these things, as well as differences in the skill and smoothness of motor coordinations. Motor skills after early infancy are very largely determined by practice. Furthermore, there is great specificity in skills. Evidently each motor function must be practiced in order for it to be performed with skill. Ability to catch a ball cannot be used to predict ability at the high jump or the broad jump. Bayley (1951a), for example, found that for ages 4½ to 12 years, scores on a battery of ten tests of manual dexterity were unrelated to strength and showed correlations of about .40 with scores on jumping (sum of three tests) and of .28 with scores on balance (sum of five tests). Scores on jumping, balance, and strength tended to correlate with each other at around .30 for most ages. Similarly, Espenschade (1940) found for children of 13 to 17 years no relation between gross and fine motor skills, though similar gross motor activities were moderately related. For example, scores on the dash usually correlated near .60 with those for the broad jump, near .40 with the jump-and-reach, and near .40 with the distance throw.

**Consistency over time.** Correlations showing the degree to which scores are consistent over time, on total motor tests, are only moderate in young children. For example, in the Berkeley growth study Bayley (1935) found that correlations of scores at 27 and 30 months with scores at six younger age levels in the first two years are, with two exceptions, below .40. At later ages, between



4 and 12 years, these Berkeley children's scores on the manual dexterity tests again show only moderate consistency over time. Scores on the 1½ year test showed correlations of .29 with scores at 4½ years, .50 at 5½ years, .49 at 6½, .55 at 7½, and .49 at 8½ years. Espenshade (1940) gave a series of motor tests to 160 children tested at six-month intervals between the ages of 13 and 17 years. She found a fair degree of consistency over a four-year period for most individual children. Glassow and Kruse (1960) found similar stability in relative scores for girls aged 6 to 14 years, with the running and jumping scores more stable than scores for throwing. Inconsistency in these cases may be attributed to the fact that scores in adolescents tend to be related to the degree of physical maturity and strength.

As measured by scores on standard tests of motor abilities, motor skills are seen to increase continuously through infancy and childhood. The increases are greatest in the first 18 months, after which the rate appears to decelerate gradually (Bayley 1951a).

**Sex differences.** There is no sex difference in motor-test scores during the first 12 years (Bayley 1939). However, after this age the girls' scores tend to stabilize while the boys' scores continue to increase (Espenshade 1947). This continued increase in boys' scores is correlated with their continuing growth in strength (Govatos 1959). Furthermore, both strength and scores in gross motor abilities are correlated in boys with their degree of physical maturation (Jones 1944; Clarke & Harrison 1962). Those who are accelerated in pubertal development are stronger and more skilled in gross motor coordinations than their more slowly maturing age peers. It is also true that the more muscular boys, with strongly masculine physiques, are stronger than those with less masculine builds (Bayley 1951b). [See INDIVIDUAL DIFFERENCES, article on SEX DIFFERENCES.]

**Specificity.** In general, after the first 15 months of age, motor skills evidence much specificity (Bayley 1951b; Espenshade 1947; Lotter 1961). It is evident also that motor skills are very responsive to practice and training (Clarke & Henry 1961; Clarke & Petersen 1961). This appears to be evident even in the very young (Holt 1960). Within normal limits and the limits of muscular strength, it should be possible to increase specific motor skills considerably through practice.

NANCY BAYLEY

[See also DEVELOPMENTAL PSYCHOLOGY; INFANCY; SENSES. Related material on child development may

be found in INTELLECTUAL DEVELOPMENT; LANGUAGE, article on LANGUAGE DEVELOPMENT; MORAL DEVELOPMENT; PERSONALITY, article on PERSONALITY DEVELOPMENT.]

#### BIBLIOGRAPHY

- BAYLEY, NANCY 1933 *Mental Growth During the First Three Years: A Developmental Study of Sixty-one Children by Repeated Tests*. *Genetic Psychology Monographs* 14, no. 1.
- BAYLEY, NANCY 1935 *The Development of Motor Abilities During the First Three Years*. Washington: Society for Research in Child Development.
- BAYLEY, NANCY 1939 *Mental and Motor Development From Two to Twelve Years*. *Review of Educational Research* 9:18-37, 114-125.
- BAYLEY, NANCY 1951a *Development and Maturation*. Pages 145-199 in Harry Helson (editor), *Theoretical Foundations of Psychology*. Princeton, N.J.: Van Nostrand.
- BAYLEY, NANCY 1951b *Some Psychological Correlates of Somatic Androgyny*. *Child Development* 22:47-60.
- BAYLEY, NANCY; and ESPENSCHADE, ANNA 1944 *Motor Development From Birth to Maturity*. *Review of Educational Research* 14:381-389.
- BAYLEY, NANCY; and ESPENSCHADE, ANNA 1950 *Motor Development and Decline*. *Review of Educational Research* 20:367-374.
- BERLYNE, D. E. 1958 *The Influence of the Albedo and Complexity of Stimuli on Visual Fixation in the Infant*. *British Journal of Psychology* 49:315-318.
- BROWN, C. A. 1961 *The Development of Visual Capacity in the Infant and Young Child*. *Cerebral Palsy Bulletin* 3:364-372.
- CARMICHAEL, LEONARD (editor) (1946) 1954 *Manual of Child Psychology*. 2d ed. New York: Wiley. → See especially pages 60-185, "The Onset and Early Development of Behavior."
- CLARKE, DAVID H.; and HENRY, FRANKLIN M. 1961 *Neuro-motor Specificity and Increased Speed From Strength Development*. American Association for Health, Physical Education, and Recreation, *Research Quarterly* 32:315-325.
- CLARKE, H. HARRISON; and HARRISON, JAMES C. E. 1962 *Differences in Physical and Motor Traits Between Boys of Advanced, Normal and Retarded Maturity*. American Association for Health, Physical Education, and Recreation, *Research Quarterly* 33:13-25.
- CLARKE, H. HARRISON, and PETERSEN, KAY H. 1961 *Contrast of Maturation, Structural, and Strength Characteristics of Athletes and Nonathletes 10 to 15 Years of Age*. American Association for Health, Physical Education, and Recreation, *Research Quarterly* 32:163-176.
- CONEL, JESSE L. 1939-1963 *The Postnatal Development of the Human Cerebral Cortex*. 7 vols. Cambridge, Mass.: Harvard Univ. Press.
- DORIS, JOHN, and COOPER, LOWELL 1964 *Brightness Discrimination in Infancy*. *American Psychologist* 19: 456 only.
- EICHORN, DOROTHY H. 1963 *Biological Correlates of Behavior*. Volume 62, pages 4-61 in National Society for the Study of Education, *Yearbook*. Part 1: Child Psychology. Univ. of Chicago Press.
- EICHORN, DOROTHY H.; and JONES, HAROLD E. 1958 *Maturation and Behavior*. Pages 211-248 in Georgene H. Seward and John P. Seward (editors), *Current*

*Psychological Issues: Essays in Honor of Robert S. Woodworth.* New York: Holt.

- ESPENSCHADE, ANNA 1940 Motor Performance in Adolescence. Society for Research in Child Development, *Monographs* 5, no. 1.
- ESPENSCHADE, ANNA 1947 Motor Development. *Review of Educational Research* 17:354-361.
- FANTZ, ROBERT L. 1958 Pattern Vision in Young Infants. *Psychological Record* 8:43-47.
- FANTZ, ROBERT L.; and ORDY, J. M. 1959 A Visual Acuity Test for Infants Under Six Months of Age. *Psychological Record* 9:159-164.
- GHEHT, LILA 1961 Developmental Changes in Tactual Thresholds on Dominant and Non-dominant Sides. *Journal of Comparative and Physiological Psychology* 54:670-673.
- GLASSOW, RUTH B.; and KRUSE, PAULINE 1960 Motor Performance of Girls Age 6 to 14 Years. American Association for Health, Physical Education, and Recreation, *Research Quarterly* 31:426-433.
- GOVATOS, LOUIS A. 1959 Relationships and Age Differences in Growth Measures and Motor Skills. *Child Development* 30:333-340.
- GRUNER, JEAN E. 1962 Histological Study of the Maturation of the Nervous System. *Developmental Medicine and Child Neurology* 4:626-639.
- HAYNES, HAROLD; WHITE, BURTON L.; and HELD, RICHARD 1965 Visual Accommodation in Human Infants. *Science* 148:528-530.
- HOLT, K. S. 1960 Early Motor Development: Posturally Induced Variations. *Journal of Pediatrics* 57:571-575.
- HOOKE, DAVENPORT 1943 Reflex Activities in the Human Fetus. Pages 17-28 in Roger Barker, Jacob S. Kounin, and Herbert F. Wright (editors), *Child Behavior and Development: A Course of Representative Studies*. New York: McGraw-Hill.
- JONES, HAROLD E. 1944 The Development of Physical Abilities. Volume 43, pages 100-122 in National Society for the Study of Education, *Yearbook*. Part 1: Adolescence. Univ. of Chicago Press.
- KEENEY, ARTHUR H. 1951 *Chronology of Ophthalmic Development: An Outline Summary of the Anatomical and Functional Development of the Visual Mechanism Before and After Birth*. Springfield, Ill.: Thomas.
- LIPSITT, LEWIS P.; and LEVY, NISSIM 1959 Electrotactual Threshold in the Neonate. *Child Development* 30:547-554.
- LOTTER, WILLARD S. 1961 Specificity or Generality of Speed of Systematically Related Movements. American Association for Health, Physical Education, and Recreation, *Research Quarterly* 32:55-62.
- SCHLUDERMANN, E.; and ZUBEK, JOHN P. 1962 Effect of Age on Pain Sensitivity. *Perceptual and Motor Skills* 14:295-301.
- SOMMERS, RONALD K.; MEYER, WILLIAM J.; and FENTON, ANN K. 1961 Pitch Discrimination and Articulation. *Journal of Speech and Hearing Research* 4:56-60.
- WEVER, ERNEST G. 1949 *Theory of Hearing*. New York: Wiley.
- WHITE, BURTON L.; CASTLE, PETER; and HELD, RICHARD 1964 Observations on the Development of Visually-directed Reaching. *Child Development* 35:349-364.

## SENSORY DEPRIVATION

See PERCEPTION, article on PERCEPTUAL DEPRIVATION.

## SENTIMENTS

See ATTITUDES; VALUES.

## SEPARATION

See FAMILY, article on DISORGANIZATION AND DISSOLUTION.

## SEPARATION OF POWERS

See CONSTITUTIONAL LAW, article on DISTRIBUTION OF POWERS; FEDERALISM; PRESIDENTIAL GOVERNMENT.

## SEQUENTIAL ANALYSIS

Sequential analysis is the branch of statistics concerned with investigations in which the decision whether or not to stop at any stage depends on the observations previously made. The motivation for most sequential investigations is that when the ends achieved are measured against the costs incurred (including the cost of making observations), sequential designs are typically more efficient than nonsequential designs; some disadvantages of the sequential approach are discussed later.

The term "sequential" is occasionally extended to cover also investigations in which various aspects of the design may be changed according to the observations made. For example, preliminary experience in an experiment may suggest changes in the treatments being compared; in a social survey a small pilot survey may lead to modifications in the design of the main investigation. In this article attention will be restricted mainly to the usual situation in which termination of a single investigation is the point at issue.

In a sequential investigation observations must be examined either one by one as they are collected or at certain stages during collection. A sequential procedure might be desirable for various reasons. The investigator might wish to have an up-to-date record at any stage, either for general information or because the appropriate sample size depends on quantities that he can estimate only from the data themselves. Alternatively, he may have no intrinsic interest in the intermediate results but may be able to achieve economy in sample size by taking them into account. Three examples will illustrate these points:

(1) An investigator may wish to estimate to within 10 per cent the mean weekly expenditure on tobacco per household. In order to determine the sample size he would need an estimate of the variability of the expenditure from household to house-



hold, and this might be obtainable only from the survey itself.

(2) A physician wishing to compare the effects of two drugs in the treatment of some disease may wish to stop the investigation if at some stage a convincing difference can already be demonstrated using the available data.

(3) A manufacturer carrying out inspection of batches of some product may be able to pass most of his batches with little inspection but may carry out further inspection of batches of doubtful quality. A given degree of discrimination between good and bad batches could be achieved in various ways, but a sequential scheme will often be more economical than one in which a sample of constant size is taken from each batch [see QUALITY CONTROL, STATISTICAL for further discussion of such applications].

The most appropriate design and method of analysis of a sequential investigation depend on the purpose of the investigation. The statistical formulation of that purpose may take one of a number of forms, usually either estimation of some quantity to a given degree of precision or testing a hypothesis with given size and given power against a given alternative hypothesis. Economy in number of observations is typically important for sequential design. Some details of particular methods are given in later sections.

Sometimes a sequential investigation, although desirable, may not be practicable. To make effective sequential use of observations they must become available without too great a delay. It would not be possible, for example, to do a sequential analysis of the effect of some social or medical policy if this effect could not be assessed until five years had elapsed. In other situations it may be possible to scrutinize the results as they are obtained, but only at very great cost. An example might be a social survey in which data could be collected rather quickly but in which a full analysis would be long and costly.

**History.** An important precursor of the modern theory of sequential analysis was the work done in 1929 by Dodge and Romig (1929-1941) on double sampling schemes. Their problem was to specify sampling inspection schemes that discriminated between batches of good and bad quality. The first stage of sampling would always be used, but the second stage would be used only if the results of the first were equivocal; furthermore, the size of the second sample and the acceptance criteria might depend on the first stage results. Bartky (1943) generalized this idea in his "multiple sampling," which allowed many stages, and his procedure

was very closely related to a particular case of the general theory of sequential analysis that Wald was developing simultaneously.

This theory, developed for the testing of military equipment during World War II, is summarized in Wald's book *Sequential Analysis* (1947). It represents a powerful exploitation of a single concept, the "sequential probability ratio test," which has provided the basis of most subsequent work. Related work proceeding simultaneously in Great Britain is summarized by Barnard (1946). Whereas Wald's theory provided the specification of a sampling scheme satisfying given requirements, Barnard's work was devoted to the converse problem of examining the properties of a given sequential scheme. Barnard drew attention to the close analogy between sequential inspection schemes and games of chance. Indeed, some of the solutions to gaming problems provided by seventeenth-century and eighteenth-century mathematicians are directly applicable to modern sequential schemes.

Postwar theoretical development, stimulated primarily by Wald's work, has perhaps outrun practical applications. Many recent workers have apparently felt that the standard sequential theory does not provide answers to the right questions, and a number of new lines of approach have been attempted.

**Sequential estimation.** Suppose that in a large population a proportion,  $p$ , of individuals show some characteristic (or are "marked") and that in a random sample of size  $n$  the number of marked individuals is  $X$ . Then the proportion of marked individuals is  $X/n$ . By standard binomial distribution theory, the standard error of  $X/n$  is  $\sqrt{p(1-p)/n}$ . The standard error expressed as a proportion of the true value is therefore  $\sqrt{(1-p)/np}$ , and when  $p$  is small this will be approximately  $(np)^{-1/2}$ . Now  $np$  is the mean value of  $X = n(X/n)$ . Intuitively, therefore, one could achieve an approximately constant proportional standard error by choosing a fixed value of  $X$ . That is, sampling would be continued until a predetermined number of marked individuals had been found. This is called a "stopping rule." The sample size,  $n$ , would be a random variable. If  $p$  happened to be very small,  $n$  would tend to be very large, an increase in  $p$  would tend to cause a decrease in  $n$ . This procedure is called "inverse sampling," and its properties were first examined by Haldane (1945).

At first sight it seems natural to estimate  $p$  by  $X/n$ . This estimator is slightly biased under inverse sampling, and some statisticians would use the modified estimator  $(X-1)/(n-1)$ , which is unbiased. Others feel that  $X/n$  is preferable despite

the bias. In most practical work the difference is negligible.

Inverse sampling is one of the simplest methods of sequential estimation. One could define different stopping rules and for any particular stopping rule examine the way in which the precision of  $X/n$  varied with  $p$ ; conversely, one could specify this relationship and ask what stopping rule would satisfy the requirement.

If a random variable is normally distributed with mean  $\mu$  and variance  $\sigma^2$ , a natural requirement might be to estimate  $\mu$  with a confidence interval of not greater than a given length at a certain probability level. With samples of fixed size, the length of the confidence interval depends on  $\sigma$ , which is typically unknown. The usual Student  $t$  procedure provides intervals of random and unbounded length. Stein (1945) describes a two-sample procedure (with preassigned confidence-interval length) in which the first sample provides an estimate of  $\sigma$ . This estimate then determines the size of the second sample; occasionally a second sample is not needed. This scheme leads naturally to a general approach to sequential estimation. Suppose that, as in inverse sampling, one proceeds in a fully sequential manner, taking one observation at a time, and stops when the desired level of precision is reached. This precision may be determined by customary standard error formulas. Anscombe (1953) showed that in large samples this procedure will indeed yield estimates of the required level of precision. Thus, suppose an investigator wished to estimate the mean number of persons per household in a certain area, with a standard error of 0.1 person, and he had little initial evidence about the variance of household size. He could sample the households until the standard error of the mean, given by the usual formula  $s/\sqrt{n}$ , fell as low as 0.1. A practical difficulty might be that of ensuring that the sampling was random.

**Sequential hypothesis testing.** Suppose that one wishes to test a specific hypothesis,  $H_0$ , in such a way that if  $H_0$  is indeed true it will usually be accepted and that if an alternative hypothesis,  $H_1$ , is true  $H_0$  will usually be rejected. In the most elementary case  $H_0$  and  $H_1$  are *simple* hypotheses; that is, each specifies completely the probability distribution of the generic random variable,  $X$ . Suppose  $f_0(x)$  and  $f_1(x)$  are the probabilities (or probability densities) that  $X$  takes the value  $x$  when  $H_0$  and  $H_1$  are true, respectively.

**Sequential probability ratio test (SPRT).** Wald proposed the following method of sequential hypothesis testing. Independent observations,  $X_i$ , are taken sequentially and result in values,  $x_i$ . Define

two positive constants,  $A_0$  and  $A_1$ . At the  $n$ th stage, calculate

$$\frac{f_{1n}}{f_{0n}} = \frac{f_1(x_1) \cdots f_1(x_n)}{f_0(x_1) \cdots f_0(x_n)}.$$

If  $f_{1n}/f_{0n} \leq A_0$ , accept  $H_0$ ; if  $f_{1n}/f_{0n} \geq A_1$ , reject  $H_0$ . If  $A_0 < f_{1n}/f_{0n} < A_1$ , take the next observation and repeat the procedure.

Wald called this procedure the "sequential probability ratio test" (SPRT). The ratio  $f_{1n}/f_{0n}$ , normally called the "likelihood ratio," plays an important part in the Neyman-Pearson theory of hypothesis testing, a fact that probably largely explains Wald's motivation. The likelihood ratio also occurs naturally in Bayesian inference [see BAYESIAN INFERENCE].

Let  $\alpha$  be the probability of rejecting  $H_0$  when it is true, and let  $\beta$  be the probability of accepting  $H_0$  when  $H_1$  is true. It can be shown that  $A_0 \geq \beta/(1-\alpha)$  and  $A_1 \leq (1-\beta)/\alpha$  and that the SPRT with  $A_0 = \beta/(1-\alpha)$  and  $A_1 = (1-\beta)/\alpha$  will usually have error probabilities  $\alpha'$  and  $\beta'$  rather close to  $\alpha$  and  $\beta$ . (The inequalities arise because sampling usually stops when the bounds  $A_0$  and  $A_1$  are slightly exceeded rather than equaled.)

The number of observations,  $n$ , required before a decision is reached is a random variable. Wald gave approximate formulas for  $E_0(n)$  and  $E_1(n)$ , the mean number of observations when  $H_0$  or  $H_1$  is true. Wald conjectured, and Wald and Wolfowitz (1948) proved, that no other test (sequential or not) having error probabilities equal to  $\alpha'$  and  $\beta'$  can have lower values for either  $E_0(n)$  or  $E_1(n)$  than those of the SPRT.

As an example, suppose that  $H_0$  specifies that the proportion,  $p$ , of "marked" individuals in a large population is  $p_0$  and that  $H_1$  states that  $p$  is  $p_1$ , where  $p_1 > p_0$ . At the  $n$ th stage, if  $x$  marked individuals have been found, the likelihood ratio is

$$\frac{f_{1n}}{f_{0n}} = \frac{p_1^x (1-p_1)^{n-x}}{p_0^x (1-p_0)^{n-x}}.$$

Sampling will continue as long as

$$\frac{\beta}{1-\alpha} < \frac{p_1^x (1-p_1)^{n-x}}{p_0^x (1-p_0)^{n-x}} < \frac{1-\beta}{\alpha},$$

using the appropriate formulas for the bounds. Taking logarithms, this inequality is

$$\begin{aligned} \log\left(\frac{\beta}{1-\alpha}\right) &< x \log\left(\frac{p_1}{p_0}\right) + (n-x) \log\left(\frac{1-p_1}{1-p_0}\right) \\ &< \log\left(\frac{1-\beta}{\alpha}\right) \end{aligned}$$



or  $a_0 + bn < x < a_1 + bn$ ,

where  $a_0$ ,  $a_1$ , and  $b$  are functions of  $p_0$ ,  $p_1$ ,  $\alpha$ , and  $\beta$ . The SPRT can thus be performed as a simple graphical procedure, with coordinate axes for  $x$  and  $n$  and two parallel boundary lines,  $x = a_0 + bn$  and  $x = a_1 + bn$  (see Figure 1). The successive

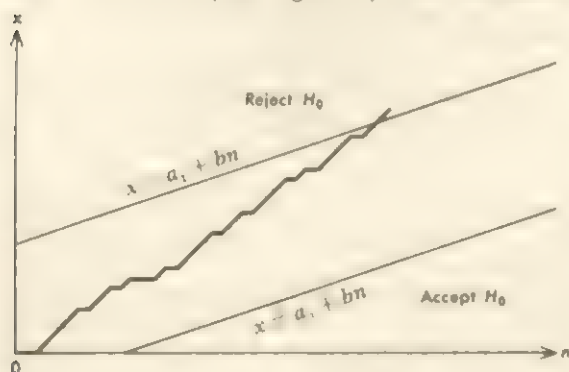


Figure 1 — Sequential probability ratio test for binomial sampling

values of  $X$  are plotted to form a "sample path," and sampling stops when the sample path crosses either of the boundaries.

Graphic solutions with parallel lines are also obtained for the test of the mean of a normal distribution with known variance and for the parameter of a Poisson distribution.

The SPRT is clearly a powerful and satisfying procedure in situations where one of two simple

hypotheses is true and where the mean number of observations is an appropriate measure of the sampling effort. Unfortunately, some or all of these conditions may not hold. A continuous range of hypotheses must usually be considered; the hypotheses may not be simple; and the variability of the number of observations, as well as its mean, may be important.

Suppose that there is a single parameter,  $\theta$ , describing the distribution of interest, and further suppose that  $H_0$  and  $H_1$  specify two particular values of  $\theta$ :  $\theta_0$  and  $\theta_1$ . For every value of  $\theta$ , including  $\theta_0$  and  $\theta_1$ , quantities of interest are the probability,  $L(\theta)$ , of accepting  $H_0$  (called the "operating characteristic" or O.C.) and the average number of observations,  $E_s(n)$  (called the "average sample number," ASN, rather than the "average sample size," for obvious reasons). Approximate formulas for both these quantities are found in Wald's book (1947). The O.C. is an approximately smooth function between 0 and 1, taking the values  $L(\theta_0) = 1 - \alpha$ ,  $L(\theta_1) = \beta$  (see Figure 2). The ASN, that is,  $E_s(n)$ , normally has a maximum for a value of  $\theta$  between  $\theta_0$  and  $\theta_1$  (see Figure 3). In the binomial problem discussed above, for example, the maximum ASN occurs close to the value  $p = b$ , for which, on the average, the sample path tends to move parallel to the boundaries. It is remarkable, though, that in many situations of practical interest this maximum value of the ASN is less than the size of the nonsequential procedure that tests  $H_0$ .

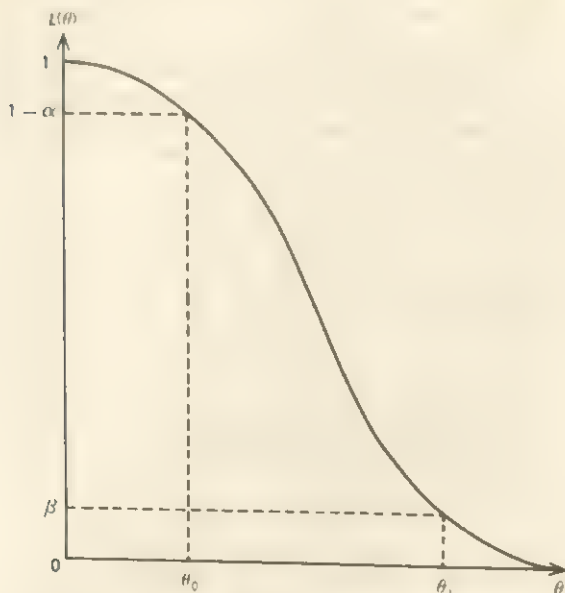


Figure 2 — Typical form of operating characteristic (O.C.)

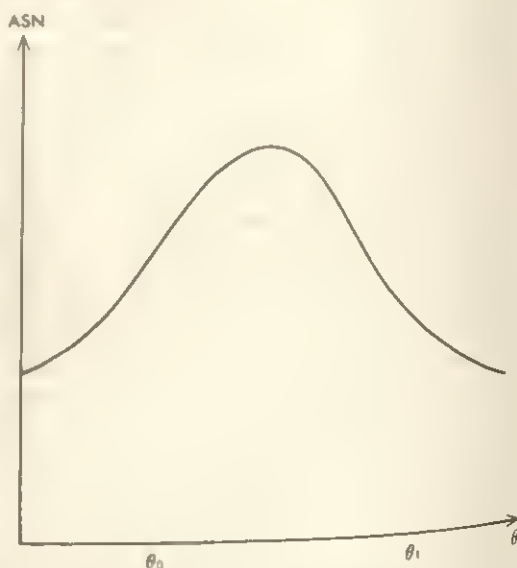


Figure 3 — Typical form of average sample number curve

and  $H_1$  with the same error probabilities as the SPRT.

*Closed procedures.* It can be shown in most cases that an SPRT, although defined for indefinitely large  $n$ , must stop some time. Individual sample numbers may, however, be very large and the great variability in sample number from one sample to another may be a serious drawback. Wald suggested that the schemes should be "truncated" by taking the most appropriate decision if a boundary had not been reached after some arbitrary large number of readings. The properties of the SPRT are, however, affected unless the truncation sample size is very high, and a number of authors (Bross 1952; Armitage 1957; Anderson 1960; Schneiderman & Armitage 1962) have recently examined closed procedures (that is, procedures with an upper bound to the number of observations) of a radically different type. In general, it seems possible to find closed procedures which are only slightly less efficient than the corresponding SPRT at  $H_0$  and  $H_1$  but which are more efficient in intermediate situations. (For an account of the use of closed procedures in trials of medical treatments, see Armitage 1960.) Frequently the hypotheses to be tested will be composite rather than simple, in that the probability distribution of the observations is not completely specified. For example, in testing the mean of a normal distribution the variance may remain unspecified. Wald's approach to this problem was not altogether satisfactory and recent work (for example, following Cox 1952) has tended to develop analogues of the SPRT using sufficient statistics where possible [see SUFFICIENCY]. A sequential  $t$ -test of this type was tabulated by Arnold (see U.S. National Bureau of Standards 1951). The usual approximations to the O.C. and ASN do not apply directly in these situations. Cox (1963) has described a large-sample approach based on maximum likelihood estimates of the parameters.

*Two-sided tests of hypotheses.* In many statistical problems two-sided tests of hypotheses are more appropriate than one-sided tests. In an experiment to compare two treatments, for example, it is customary to specify a null hypothesis that no effective difference exists and to be prepared to reject the hypothesis if differences in either direction are demonstrable. One approach to two-sided sequential tests (used, for example, in the standard sequential  $t$ -test) is to allow an alternative composite hypothesis to embrace simple hypotheses on both sides of the null hypothesis. It may be more appropriate to recognize here a three-decision problem, the decisions being to accept the null hypothesis

(that is, to assert no demonstrable difference), to reject it in favor of an alternative in one direction, or to reject it in the other direction. A useful device then is to run simultaneously two separate two-decision procedures, one to test  $H_0$  against  $H_1$  and the other to test  $H_0$  against  $H'_1$  where  $H_1$  and  $H'_1$  are alternatives on different sides of  $H_0$  (see Sobel & Wald 1949).

*Other approaches.* Wald's theory and the sort of developments described above are in the tradition of the Neyman-Pearson theory of hypothesis testing, with its emphasis on risks of accepting or rejecting certain hypotheses when these or other hypotheses are true. This approach is arbitrary in many respects; for example, in the SPRT there is no clear way of choosing values of  $\alpha$  and  $\beta$  or of specifying an alternative to a null hypothesis. Much theoretical work is now based on statistical decision theory [see DECISION THEORY; see also Wald 1950]. The aim here is to regard the end product of a statistical analysis as a decision of some sort, to measure the gains or losses that accrue, under various circumstances, when certain decisions are taken, to measure in the same units the cost of making observations, and to choose a rule of procedure that in some sense leads to the highest expectation of gain. Prior probabilities may or may not be attached to various hypotheses. With certain assumptions, the SPRT emerges as an optimal solution for the comparison of two simple hypotheses, but there is no reason to accept it as a general method of sequential analysis. Chernoff (1959) has developed a large-sample theory of the sequential design of experiments for testing composite hypotheses. The aim is to minimize cost in the limiting situation in which costs of wrong decisions far outweigh costs of experimentation. Account is taken of the choice between different types of observation (for example, the use of either of two treatments). A somewhat different approach (for instance, Wetherill 1961) is to stop an investigation as soon as the expected gain achieved by taking a further observation is outweighed by the cost of the observation. The formulation of the problem requires the specification of prior probabilities and its solution involves dynamic programming [see PROGRAMMING].

The recent interest shown in statistical inference by likelihood, with or without prior probabilities, has revealed a conflict between this approach and the more traditional methodology of statistics, involving significance tests and confidence intervals [see LIKELIHOOD]. In the likelihood approach, inferences do not depend on stopping rules. There is, on this view, no need to have a separate theory of estima-



tion for inverse sequential sampling or to require that investigations should follow a clearly defined stopping rule before their results can be rigorously interpreted.

The effect of all this work is at present hard to assess. The attraction of a global view is undeniable. On the other hand, the specification of prior distributions and losses may be prohibitively difficult in most scientific investigations.

A number of workers have studied problems involving progressive changes in experimental conditions. Robbins and Monro (1951) give a method for successive approximation to the value of an independent variable, in a regression equation, corresponding to a specified mean value of the dependent variable. Similar methods for use in stimulus-response experiments are reviewed by Wetherill (1963). In industrial statistics much attention has been given to the problem of estimating, by a sequence of experiments, the set of operating conditions giving optimal response. This work has been stimulated mainly by G. E. P. Box, whose *evolutionary operation* is a method by which continuous adjustments to operating conditions can be made [see EXPERIMENTAL DESIGN, article on RESPONSE SURFACES].

P. ARMITAGE

[Directly related are the entries ESTIMATION; HYPOTHESIS TESTING; SCREENING AND SELECTION.]

#### BIBLIOGRAPHY

Useful bibliographies and surveys of sequential analysis are given in Jackson 1960; Johnson 1961, and Wetherill 1966. Wald 1947 remains an important source book; an elementary exposition of the SPRT, with examples, is contained in Columbia University, Statistical Research Group 1945. Armitage 1960 is concerned with medical applications. Federer 1963 discusses sequential procedures in screening and selection problems.

ANDERSON, T. W. 1960 A Modification of the Sequential Probability Ratio Test to Reduce the Sample Size. *Annals of Mathematical Statistics* 31:165-197.

ANScombe, F. J. 1953 Sequential Estimation. *Journal of the Royal Statistical Society Series B* 15:1-29. → Contains eight pages of discussion.

ARMITAGE, P. 1957 Restricted Sequential Procedures. *Biometrika* 44:9-26.

ARMITAGE, P. 1960 *Sequential Medical Trials*. Oxford: Blackwell; Springfield, Ill.: Thomas.

BARNARD, G. A. 1946 Sequential Tests in Industrial Statistics. *Journal of the Royal Statistical Society Supplement* 8:1-26. → Contains four pages of discussion.

BARTKY, WALTER 1943 Multiple Sampling With Constant Probability. *Annals of Mathematical Statistics* 14:363-377.

BROSS, I. 1952 Sequential Medical Plans. *Biometrics* 8:188-205.

CHERNOFF, HERMAN 1959 Sequential Design of Experiments. *Annals of Mathematical Statistics* 30:755-770.

COLUMBIA UNIVERSITY, STATISTICAL RESEARCH GROUP 1945 *Sequential Analysis of Statistical Data: Applications*. New York: Columbia Univ. Press.

COX, D. R. 1952 Sequential Tests for Composite Hypotheses. *Cambridge Philosophical Society, Proceedings* 48:290-299.

COX, D. R. 1963 Large Sample Sequential Tests for Composite Hypotheses. *Sankhyā: The Indian Journal of Statistics* 25 5-12.

DODGE, HAROLD F.; and ROMIG, HARRY G. (1929-1941) 1959 *Sampling Inspection Tables: Single and Double Sampling*. 2d ed., rev. & enl. New York: Wiley; London: Chapman. → This book is a republication of fundamental papers published by the authors in 1929 and in 1941 in the *Bell System Technical Journal*.

FEDERER, WALTER T. 1963 Procedures and Designs Useful for Screening Material in Selection and Allocation, With a Bibliography. *Biometrics* 19:553-587.

HALDANE, J. B. S. 1945 On a Method of Estimating Frequencies. *Biometrika* 33 222-225.

JACKSON, J. EDWARD 1960 Bibliography on Sequential Analysis. *Journal of the American Statistical Association* 55:561-580.

JOHNSON, N. L. 1961 Sequential Analysis: A Survey. *Journal of the Royal Statistical Society Series A* 124:372-411.

ROBBINS, HERBERT; and MONRO, SUTTON 1951 A Stochastic Approximation Method. *Annals of Mathematical Statistics* 22 400-407.

SCHNEIDERMAN, M. A.; and ARMITAGE, P. 1962 A Family of Closed Sequential Procedures. *Biometrika* 49:41-56.

SOBEL, MILTON; and WALD, ABRAHAM 1949 A Sequential Decision Procedure for Choosing One of Three Hypotheses Concerning the Unknown Mean of a Normal Distribution. *Annals of Mathematical Statistics* 20 502-522.

STEIN, CHARLES M. 1945 A Two-sample Test for a Linear Hypothesis Whose Power is Independent of the Variance. *Annals of Mathematical Statistics* 16:243-258.

U.S. NATIONAL BUREAU OF STANDARDS 1951 *Tables to Facilitate Sequential t-Tests*. Edited by K. J. Arnold. National Bureau of Standards, Applied Mathematics Series, No. 7. Washington: U.S. Department of Commerce.

WALD, ABRAHAM 1947 *Sequential Analysis*. New York: Wiley.

WALD, ABRAHAM (1950) 1964 *Statistical Decision Functions*. New York: Wiley.

WALD, ABRAHAM, and WOLFOWITZ, J. (1948) 1955 Optimum Character of the Sequential Probability Ratio Test. Pages 521-534 in Abraham Wald, *Selected Papers in Statistics and Probability*. New York: McGraw-Hill. → First published in Volume 19 of the *Annals of Mathematical Statistics*.

WETHERILL, G. B. 1961 Bayesian Sequential Analysis. *Biometrika* 48:281-292.

WETHERILL, G. B. 1963 Sequential Estimation of Quantal Response Curves. *Journal of the Royal Statistical Society Series B* 25:1-48. → Contains nine pages of discussion.

WETHERILL, G. BARRIE 1966 *Sequential Methods in Statistics*. New York: Wiley.

#### SERIAL CORRELATION

See TIME SERIES.

## SERRA, ANTONIO

Antonio Serra was an Italian economist born around the middle of the sixteenth century in Cosenza, a city of Calabria then belonging to the Kingdom of Naples. Nothing is known about his life except that he spent some time in a Neapolitan prison, charged—according to different sources—either with forging money or with participating in a conspiracy, headed by his fellow citizen Tommaso Campanella, to overthrow the Spanish government; it was in prison that he wrote his *Breve trattato delle cause che possono far abbondare li regni d'oro e d'argento, dove non sono miniere, con applicazione al regno di Napoli* (1613; "A Brief Treatise on the Causes That Can Make Gold and Silver Plentiful in Kingdoms Where There Are No Mines, With Application to the Kingdom of Naples"). Serra may have written *Breve trattato* and dedicated it to the Spanish viceroy in the hope of regaining his freedom. In 1617, while still in prison, he obtained permission to submit to the court his ideas for economic reform; but his plan was not accepted and he was sent back to prison. Serra's work remained practically unknown until Ferdinando Galiani extolled it in one of the footnotes to the 1780 edition of his *Della moneta* ([1750] 1915, pp. 344–345).

Serra refuted the view expressed by Marc'Antonio de Santis, author of *Discorso intorno alli effetti che fa il cambio in regno* (1605; "Discourse on the Effects of the Rate of Exchange in the Kingdom"), that the scarcity of money in the Kingdom of Naples was attributable to the high rate of exchange. He used arguments which made him the first to elaborate the view that the rate of exchange is a consequence of the balance of international payments, which alone regulates the international flow of money (Tagliacozzo 1937, p. xxx; Schumpeter 1954, p. 354).

Serra had a clear understanding of the balance-of-trade concept as an analytic tool in economic reasoning and was the first to use that tool fully and, on the whole, correctly. [See INTERNATIONAL MONETARY ECONOMICS, article on BALANCE OF PAYMENTS.] In analyzing the component items of the balance of payments, he paid due attention to the invisible items; in this respect he anticipated and surpassed all other seventeenth-century writers, including the overrated Thomas Mun, whose *England's Treasure by Foreign Trade* was written about fifty years after the *Breve trattato* and may have been influenced by it (Mazzei 1924, p. 396; Schumpeter 1954, pp. 353–354; Tagliacozzo 1937, pp. xxx–xxxiii).

Serra's book went beyond the immediate implications of its title, explaining the outflow of gold and silver from the Kingdom of Naples and the balance of trade as consequences of the economic conditions of the country.

Essentially, the treatise is about the factors on which depend the abundance not of money but of *commodities*—natural resources, quality of the people, the development of industry and trade, the efficiency of government—the implication being that if the economic process as a whole functions properly, the monetary element will take care of itself and not require any specific therapy. And this argument contains several contributions to the nascent stock of theoretical tools. (Schumpeter 1954, p. 195)

Hence, Schumpeter credits Serra "with having been the first to compose a scientific treatise, though an unsystematic one, on Economic Principles and Policy" (*ibid.*).

One of the theoretical tools contained in the *Breve trattato* (part 1, chapter 3) is the first clear formulation of the law of diminishing returns in agriculture and increasing returns in manufacturing industries [see Tagliacozzo 1937, p. xxxd; see also PRODUCTION].

Serra's realistic, active approach to economic policy—exemplified by his plea for the introduction of new industries in the Kingdom of Naples (see *Breve trattato*, part 3, chapter 8)—probably struck Galiani ([1750] 1915, pp. 344–345) as an anticipation of his own relativistic, undogmatic, pragmatic, liberal (as opposed to laissez-faire) approach. [See the biography of GALIANI; on Galiani's approach to economic policy, see Tagliacozzo 1937, pp. lii–lv.] This is what caused Friedrich List—the founder of the "infant-industry" argument and a forerunner of the historical school of economics—to say, in a chapter devoted almost entirely to Serra (1841, book 3, chapter 1 in 1856 edition), "Antonio Serra sees the nature of things as it actually exists and not through the spectacles of previous systems or of some one principle which he is determined to advocate and carry out."

GIORGIO TAGLIACOZZO

## BIBLIOGRAPHY

- ARIAS, GINO 1923 Il pensiero economico di Antonio Serra. *Politica* 16: 129–146.  
 BENINI, RODOLFO 1892 Sulle dottrine economiche di A. Serra: Appunti critici. *Giornale degli economisti* Second Series 5: 222–248.  
 FERNARI, TOMMASO 1880 Studi sopra Antonio Serra e Marc'Antonio De Santis. Milan: Hoepli.  
 FERNARI, TOMMASO 1882 Delle teorie economiche nelle provincie napoletane. Milan: Hoepli.  
 GALIANI, FERDINANDO (1750) 1915 *Della moneta*. Edited by Fausto Nicolini. New ed. Bari: Laterza. → Serra's



work was first mentioned in a footnote to the 1780 edition.

- GOBBI, ULISSE 1884 *La concorrenza estera e gli antichi economisti italiani*. Milan: Hoepli.
- GOBBI, ULISSE 1889 *L'economia politica negli scrittori italiani del sec. XVI-XVII: Studio*. Milan: Hoepli.
- GRAZIANI, AUGUSTO (editor) 1913 *Economisti del cinque e seicento*. Bari: Laterza.
- LIST, FRIEDRICH (1841) 1928 *The National System of Political Economy*. London: Longmans. → First published in German.
- MAZZEI, JACOPO 1924 *Politica economica internazionale inglese prima di Adamo Smith*. Milan: Vita e Pensiero.
- SANTIS, MARC'ANTONIO DE 1605 *Discorso intorno alli effetti che fa il cambio in regno*. Naples: Vitale.
- SCHUMPETER, JOSEPH A. (1954) 1960 *History of Economic Analysis*. Edited by E. B. Schumpeter. New York: Oxford Univ. Press.
- SERRA, ANTONIO (1613) 1913 *Breve trattato delle cause che possono far abbondare li regni d'oro e d'argento, dove non sono miniere, con applicazione al regno di Napoli*. Pages 141-235 in A. Graziani (editor), *Economisti del cinque e seicento*. Bari: Laterza. → Translated extracts appear on pages 145-167 in Arthur E. Monroe (editor), *Early Economic Thought*, published in 1924 by Harvard University Press. A full translation, with an introduction by Klaus-Peter Heiss, is projected for publication by the Princeton University Press.
- TAGLIACCOZZO, GIORGIO (editor) 1937 *Economisti napoletani dei sec. XVII e XVIII*. Bologna: Cappelli.
- VITI DE MARCO, ANTONIO DE (1891) 1898 *Le teorie economiche di Antonio Serra*. In Antonio De Viti de Marco, *Saggi di economia e finanza*. Rome: Giornale degli Economisti.

## SET

See ATTENTION; RESPONSE SET.

## SETTLEMENT

See CITY; COMMUNITY; NEIGHBORHOOD; VILLAGE.

## SEX DIFFERENCES

See under INDIVIDUAL DIFFERENCES.

## SEXUAL BEHAVIOR

- |                                              |                    |
|----------------------------------------------|--------------------|
| I. ANIMAL SEXUAL BEHAVIOR                    | Charles H. Phoenix |
| II. SOCIAL ASPECTS                           | Samuel Z. Klausner |
| III. SEXUAL DEVIATION: PSYCHOLOGICAL ASPECTS | John Money         |
| IV. SEXUAL DEVIATION: SOCIAL ASPECTS         | John H. Gagnon     |
| V. HOMOSEXUALITY                             | Evelyn Hooker      |

### I

#### ANIMAL SEXUAL BEHAVIOR

Animal sexual behavior is but one aspect of the total pattern of reproductive behavior. Sexual be-

havior includes all responses directly associated with genital stimulation and copulation, whether homosexual or heterosexual. Among warm-blooded vertebrates, heterosexual copulation culminates in the transfer of sperm from male to female, with subsequent fertilization of ova and reproduction of the species.

Despite its significance for species survival, sexual behavior has received relatively little experimental attention. With few exceptions, sexual behavior in man has not been studied experimentally and is likely to remain generally inaccessible for study by experimentalists for some time to come (Ford & Beach 1951; Lloyd 1964). Careful experimental analysis of animal sexual behavior, especially that of primates, should provide a valuable background of information for understanding the data that exist concerning sexual behavior in man. Many of the forms of sexual behavior described by Kinsey and his associates (Kinsey et al. 1948, 1953) as part of the human sexual behavior repertoire have been observed in a number of other mammalian species. Beach (1949) has suggested that a study not only of the similarities but of the differences in sexual behavior between man and other mammalian species should help in our understanding of how social forces can modify a basic biological drive.

In experimental research on sexual behavior, as in research on any other aspect of behavior, the experimentalist seeks to discover and analyze those variables of which behavior is a function. The initial step in the process is the identification of components of sexual behavior and a description of these in relation to the total pattern. It then becomes possible to measure quantitatively the latency, frequency, and duration of each component and thus to assess the effects of the independent variables on specific components of the total sexual behavior pattern.

Since there have been many reviews of sexual behavior patterns of a number of insects, fishes, amphibians, reptiles, and birds, the present account is restricted to a discussion of sexual behavior in placental mammals. Detailed descriptions are available of the mating behavior of several laboratory mammals, including the rat, guinea pig, hamster, rabbit, cat, dog, monkey, and chimpanzee, and of farm animals such as the goat, sheep, cow, pig, and horse. Limited descriptions are also available of sexual behavior patterns in normally wild species not generally found in laboratories, such as the chinchilla, porcupine, skunk, mink, fox, and elephant. The accuracy and detail of the descrip-

tions of the behavior in the last-mentioned species vary widely, however, and further study is needed.

### Components of sexual behavior

**Mounting.** The most readily identified aspect of animal sexual behavior is the mounting response. It is an essential component of the male mating pattern but frequently occurs in females, especially during estrus. In a successful mating sequence the male mounts the female from the rear. The forelegs are placed on the back or around the sides of the female with the hind legs on the ground. The male may then execute a series of rapid pelvic thrusts. The posture and motion permit the male to insert the erect penis into the female's vagina. Not all mounting, however, is oriented to the rear of the female. The male may mount the side or head of the partner and may or may not execute pelvic thrusts. Mounting by females also shows similar variation. Females of a number of species display mounting behavior only during estrus; but in some species, such as the rat and dog, mounting by females is not restricted to any particular part of the estrous cycle. The mount executed by a female is frequently indistinguishable from that of the male. Thus, the typography of the behavior does not identify the sex of the animal. Mounting occurs between members of the same or opposite sex and across species in some instances.

**Intromission.** Intromission is obviously restricted to the male. To achieve intromission, not only must the male be properly oriented, but the penis must be erect. The detection of intromission is difficult in most species and often requires special conditions of observation. In some species, such as the guinea pig, the rate of pelvic thrusting may be slower during intromission. In the rat, when intromission is not accompanied by ejaculation, the intromission response is usually followed by a vigorous backward lunge and a period during which the male may manipulate the penis with paws and mouth before mounting again. This gross bodily reaction is most frequently used as an index of intromission rather than visualization of penis insertion.

**Ejaculation.** Ejaculation, as in the case of intromission, is not identified directly but is commonly inferred by changes in gross body movement. In the guinea pig, for example, at the moment of ejaculation the flanks of the male are drawn in as in a spasm. The male then dismounts, or the female pulls away, and both male and female clean the genitalia. Ejaculation by the male rat is also accompanied by a drawing in of the flanks. Follow-

ing ejaculation the male dismounts much more slowly than he does following intromission, and there ensues a period of relative inactivity. Judging whether or not ejaculation has occurred from observation of gross behavior is particularly difficult in some species, and the presence of sperm or an ejaculatory plug in the vagina is frequently used as confirmatory evidence.

**Lower components.** Not only is mounting the most conspicuous aspect of sexual behavior but together with intromission and ejaculation shows for most species the least variation among the numerous components of the sexual behavior pattern. Wide variation exists from species to species in the so-called lower components of the mating pattern. In general, these lower components appear to produce excitement and readiness to mate in both sexes. The guinea pig displays a mating pattern typical of many species. The following sequence is characteristic: On encountering an estrous female, the male, walking with a characteristic gait, circles the female; he rubs against her flanks, nibbles at the ear, prods the female's mouth area with his nose, licks the genitalia, then approaching from the rear, rubs his nose and lower jaw over the female's back, a response that frequently elicits the lordosis response in the receptive female.

These lower components may occur in any order and do not represent a stereotyped sequence of reflexes triggered by the presence of an estrous female. In some instances, especially in experienced individuals, the male may mount the female without any preliminary display and may achieve intromission and ejaculation without manifesting any of the lower components.

**Lordosis.** The one outstanding component of female sexual behavior common to all mammalian species except man is the lordosis response. In the monkey the receptive posture is referred to as a "present." The receptive posture or lordosis consists of a particular stance, with the four legs fixed and a straightening and arching (in some cases) of the back with elevation of the pudendum. In some species the head is thrown backward and the tail is deviated or displaced so as to expose the genitalia. A number of minor species differences exist, but in all species the pattern facilitates both mounting and intromission by the male.

The complete pattern of female sexual behavior may include a number of other components in addition to the lordosis response. These additional components vary widely among species. They may precede copulation and may be classed as courtship



behavior or "foreplay" or precede the onset of full estrous behavior as in the guinea pig and be referred to as proestrous behavior. In some species, such as the cat, estrus may be followed by distinct behavioral responses, termed afterreaction.

### Factors influencing sexual behavior

Accurate description of sexual behavior in natural surroundings and under carefully controlled laboratory conditions, although an essential first step, constitutes but a starting point for the study of the major problems associated with sexual behavior. In experimental investigation, sexual behavior has been found to vary with the species, strain, sex, diet, temperature, illumination, season, age, method of rearing (including age of weaning), nature of housing (including social and isolated conditions), stage in the reproductive cycle, familiarity with the test area (territoriality) and test partner, duration and frequency of tests, dominance relationship, and the general health of the animal. With the exception of masturbation, sexual behavior is a social phenomenon; hence the behavior of each individual of a pair or group constitutes an added stimulus variable not readily controlled in the assessment of sexual behavior. It is frequently convenient to classify the determinants of sexual behavior into a few broad categories. Such categories are usually neither all-inclusive nor mutually exclusive but suggest, in a general way, the class of variable emphasized by an investigator or represented in a given research effort. For purposes of presentation we may categorize the variables as genetic, hormonal, sensory and neural, and experiential.

**Genetic variables.** Chromosomal sex is determined at the time of fertilization and depends upon the particular chromosomes contributed by the parents. Many problems of sex determination at the gene level of analysis remain to be solved, but these need not concern us here. Classification of the individual as male or female is usually based on the appearance of the external genitalia. Unfortunately sex classification based on this criterion does not always correspond to chromosomal or gonadal sex, nor does it necessarily correspond to the pattern of sexual behavior displayed by the animal. In addition to these possible inconsistencies, chromosomal sex itself may not be normal, as for example in the case of an individual whose chromosomal pattern is XO and who, accordingly, lacks one of the sex chromosomes. Nevertheless, species differences, strain differences, and sex differences may be changed by varying genetic background. Such knowledge contributes little to our under-

standing of how the genes bring about their effect; but regardless of the mechanisms involved, it is assumed that all organismic variables have a genetic basis (Beach 1947a).

**Breeding experiments.** The genetic material contributed by the parents has been shown to determine the particular pattern of sexual behavior that will be displayed by the individual as an adult. Not only is it possible to breed for high or low levels of sexual behavior, but different modes of inheritance of the various components of sexual behavior have been demonstrated. For example, in a study of two inbred strains of guinea pigs, in which the male of the two strains differed in ejaculation latency and in mounting frequency, Jakway (1959) has shown that ejaculation latency of one strain was dominant and mounting frequency of the second strain was dominant when these were studied in  $F_1$  and  $F_2$  hybrid generations and in backcrosses. Differences have also been demonstrated in modes of inheritance of components of female sexual behavior in the guinea pig (Goy & Jakway 1959). It is obvious that sexual behavior need not be inherited as a unitary trait. Although most work on the inheritance of patterns of sexual behavior has been carried out on insects, an increasing amount of work is being done on mammals (Dilger 1962; Goy & Jakway 1962).

**Hormone variables.** Sexual behavior among females of lower mammalian species is characterized by its cyclicity. The display of sexual behavior is referred to as estrus or heat. The period coincides with the presence of mature ovarian follicles and, generally, with changes in the vaginal epithelium. The term estrus is sometimes used to refer to the physiological state of the ovary or condition of the cells in the vagina. The multiple meaning attached to the term is testimony to the close relationship between the physiological state and the behavioral state. Ovulation usually, although not always, accompanies estrus, and thus mating is most likely to occur at a time when fertilization is possible. Among primates, such as the monkey, chimpanzee, and man, the display of sexual behavior is not confined to the period of ovulation but may occur at any point in the cycle (Eayrs & Glass 1962; Young 1941). The decreasing dependency on ovarian hormones has been taken as evidence that animals higher on the phylogenetic scale have through evolution become emancipated from gonadal hormone control of sexual behavior (Beach 1947b).

Females and males of some species, such as the marten and deer, are seasonal breeders. Sexual activity among seasonal breeding males is likely to be confined to periods when the females of the

species are in estrus. Among nonseasonal breeding species, males do not show the cyclic display of sexual behavior characteristic of the females. Correspondingly the level of gonadal hormone in the male does not vary in any cyclic fashion.

*Gonadectomy and replacement therapy.* Perhaps the most compelling evidence of the importance of the gonadal hormones in the display of sexual behavior comes from studies of gonadectomy and replacement therapy.

Spaying adult females of subprimate species, with the possible exception of the rabbit, results in the complete and immediate loss of mating behavior. But the injection of estrogen or estrogen and progesterone brings the female into estrus. With appropriate timing and dosage of injected hormone, estrus thus induced may be indistinguishable from that displayed prior to spaying. In primates below man, the effects of spaying are not as conspicuous as in lower mammals, and although the effectiveness of replacement therapy has not been as carefully studied, reports suggest that the dramatic change in behavior observed in lower mammals is not seen in primates. The effects of ovariectomy on human sexual behavior are not fully known, but some sexual behavior is known to persist following the operation. In general, the less conspicuous are the changes in behavior at estrus, the less effective ovariectomy appears to be in eliminating receptivity.

Castration of the adult male does not result in complete and immediate loss of sexual behavior as does spaying in the female. The first component of the mating pattern to be lost following castration is the ejaculatory response, and later intromission drops out of the repertoire. There is considerable species variation, but in many species the male may continue to mount and show pelvic thrusts months or years after castration. The decrement in performance is generally greater in rodents and lagomorphs than in carnivores (Beach 1958). Information on the effects of castration in primates is scanty. Present evidence suggests that the prepuberally castrated monkey and ape retain the capacity for erection, and they display the normal mounting pattern (Goy 1964). Little is known of effects of castration on human sexual behavior.

Just as there is a gradual diminution of sexual behavior following castration, restoration of the complete pattern of sexual behavior by androgen administration is a gradual process. An increased frequency of mounting is followed by the occurrence of intromission and finally the ejaculatory response is restored. Such, at least, is the case in the lower mammals that have been studied. No

quantitative studies exist on the effectiveness of androgen replacement therapy in subhuman primates, and the evidence for man is contradictory.

There is no question of the importance of sex hormones for the full expression of sexual behavior in adults of lower mammalian species; however, sex hormone specificity remains somewhat ambiguous. In either sex the heterotypic hormone is not as effective as the homotypic hormone in promoting the display of appropriate sexual behavior. In general, the character of the behavior induced by hormone injection, whether masculine or feminine, is not a property of the hormone itself but depends upon the organization of the sexual behavior mechanisms within the individual.

The role of gonadal hormones in differentiation of biological sex has been studied intensively by embryologists over many years (Burns 1961), but the role of these hormones in differentiation of sexual behavior has, until recently, received little attention.

It has now been demonstrated that in both the genetic female guinea pig and the rat, testosterone administration during the period of differentiation of the tissues that are to mediate sexual behavior, or what we have called sexual behavior mechanisms, results in permanent suppression of female behavior and in an increase in display of male behavior. In the guinea pig, this period of differentiation is completed prenatally, whereas in the rat the period extends into the early postnatal period. The behavior of genetic female adult rhesus monkeys treated prenatally with testosterone has not been studied, but observation of behavior during the first year of life showed a masculinization of patterns of play behavior and indications of an increase in the frequency of display of male sexual behavior (Phoenix et al. 1959; Young et al. 1964).

Further evidence indicating the significance of testosterone in differentiation of tissues mediating the pattern of sexual behavior is provided by the response of rats castrated the day of birth and injected as adults with estrogen and progesterone. These males displayed the complete female estrous response with a lordosis frequency comparable to that of females. Such was not the case when rats were castrated at ten days of age or later.

In adulthood the role of testosterone is to activate pre-established patterns of sexual behavior. During the prenatal or early postnatal period the role of testosterone is organizational, in the sense that it differentiates in both genetic males and females a prepotency for the male pattern of sexual behavior (Young et al. 1964).

The gonadal hormones are major determinants



of sexual behavior both as activators and organizers of patterns of sexual behavior. The non-gonadal hormones play a less direct role in control of sexual behavior. The pituitary, for example, derives its significance for behavior primarily because it controls gonadal action. Sexual behavior can be induced in hypophysectomized animals by injection of gonadal hormones, but in otherwise intact animals pituitary dysfunction is associated with abnormal reproductive function and sexual behavior.

**Thyroid activity.** Broad species and individual differences exist with respect to the importance of thyroid activity. The thyroid is not crucial to normal sexual behavior in the male guinea pig or rat. The thyroidectomized female rat may mate successfully but the spayed, thyroidectomized guinea pig exhibits reduced responsiveness to injected estradiol. In the bull, thyroidectomy abolishes all sexual behavior. Present evidence emphasizes species differences, and when thyroid deficiency appears to interfere with normal sexual behavior it most likely does so by altering systemic metabolic function.

**Sensory variables.** The senses serve several functions vital to sexual behavior. They are essential in bringing together the male and female of a given species and in arousing and directing behavior that eventually results in copulation and reproduction. The distance receptors play a major role in locating appropriate sex partners, and all of the senses probably contribute to the arousal and performance of the behavior sequence culminating in insemination.

The effects of sensory deprivation in sexual behavior have been studied to a limited extent in the rat, rabbit, and cat. Beach (1951) has concluded that sexual arousal in the male does not depend upon any one sense modality but upon the total pattern of sensory input from the several receptor systems. Generally the female is less dependent upon multiple stimulation for display of adequate sexual behavior. A female deprived of vision, olfaction, and audition may continue to mate successfully, but this is not true of the male.

**Neural variables.** The neural tissues constitute what is presumed to be one of the primary loci of gonadal hormone action controlling behavior. The site of action within the nervous system is not definitely known, but it is assumed to include both spinal cord and brain.

**Spinal cord.** When the cord is transected above the lumbar region in males of species ranging from rat to man, stimulation of the penis still produces erection, ejaculation, and pelvic thrusting. The majority of paraplegics can achieve erection, and some

have had fertile matings. It is obvious, therefore, that at least some of the more reflexive elements of the sexual behavior pattern are organized at the spinal level. Effects of cord transection in the female are not as clear, but whatever fragmentary elements of sexual behavior survive cord transection occur independently of hormonal stimulation. In neither sex is there anything resembling the normal pattern of sexual behavior in preparations with the brain stem sectioned below the level of the hypothalamus.

**Hypothalamus.** Small bilateral and midventral lesions in the hypothalamus of spayed female guinea pigs may eliminate all receptive behavior and mounting despite hormonal injection, which in the spayed unoperated female induces the complete estrous response (Brookhart et al. 1940). In some females with slightly different lesions, receptive behavior is blocked but mounting behavior is unaffected. In still other cases, mounting is eliminated but receptive behavior persists (Goy & Phoenix 1963). In the male guinea pig, hypothalamic lesions may eliminate all sexual behavior. However, some males may continue to mount following the operation. In the absence of testicular atrophy, loss of sexual behavior cannot be attributed to hormonal insufficiency (Brookhart & Dey 1941; Phoenix 1961).

Hypothalamic lesions in a number of species, including the rat (Soulairac & Soulairac 1956), rabbit, cat, and ewe, have resulted in comparable loss of sexual behavior. In most of these species, although not yet established for both sexes, one area in or near the hypothalamus controls pituitary function and thus governs sexual behavior indirectly; and a second area exerts direct neural control over sexual behavior. Destruction of the former eliminates sexual behavior by interfering with pituitary gonadal axis function and therefore with output of gonadal hormone necessary to normal sexual behavior. Destruction of the latter area eliminates sexual behavior by interfering with a primarily neural mechanism (Sawyer 1960).

Lesions in the anterior portion of the hypothalamus in spayed female rats and guinea pigs have produced persistent receptivity, or estrus. Such an operation releases the animal, to some extent at least, from gonadal hormone control of sexual behavior and thus produces a condition thought to exist normally in a number of primates, especially man. The operation may be interpreted as removing a neural area that in the normal state inhibits sexual behavior unless stimulated by female hormones (Goy & Phoenix 1963).

A number of other lines of evidence implicate

the hypothalamus as the region of the brain most directly involved in the control and integration of sexual behavior. For example, hypothalamic implants of solid diethylstilbestrol dibutyrate produce in the spayed cat the full display of sexual behavior. Comparable implants in other regions of the brain or in subcutaneous tissue do not produce this effect (Harris et al. 1958).

Experiments involving electrical recording from hypothalamic areas in the cat have shown changes in potential associated with estrous reactions, and changes in electroencephalograph activity of the anterior hypothalamus have also been observed in association with mating behavior in the female rabbit (Sawyer 1960).

When an electrode is permanently implanted in the hypothalamus of a rat and so wired that pressing a bar delivers a small electrical impulse to the brain area at the electrode tip, rats will repeatedly press the bar, thus obtaining electrical stimulation to the area. Castration abolishes the bar-pressing response by the rat, but the response can be reinstated by injecting the castrate with testosterone (Olds 1958). Such a demonstration tells us little about sexual behavior but implicates the hypothalamus in behavior that is dependent on a gonadal hormone.

In a now classic experiment, Klüver and Bucy (1939) performed bilateral temporal lobectomies on monkeys and reported a broad constellation of behavioral changes, including heightened sexual activity, following the operation. Hypersexuality was later reported following destruction of the amygdaloid complex and overlying pyriform cortex in lynxes, agoutis, cats, and monkeys. Castration of male cats resulted in elimination of hypersexuality. The lesion is interpreted as having had its effect by removing from inhibition the neuroendocrine mechanisms regulating the display of sexual behavior (Schreiner & Kling 1956).

**Cerebral cortex.** In lower mammals, including the cat, complete removal of the neocortex of the female does not eliminate sexual receptivity, although the quality of the response may suffer. In male rats comparable decortication eliminates all sexual behavior, and these males show no signs of arousal in the presence of an estrous female. Partial decortication in the cat may eliminate successful mating behavior because of loss of sensory-motor integration, but the operation need not interfere with sexual arousal (Beach 1951).

In primates the cortex may play a major role in integration of sexual behavior. In general, the more important learning and memory are in sexual performance, the more important the neocortex is likely

to be. It is apparent that control of sexual behavior is not vested in any single brain area. The evidence suggests rather that the entire nervous system is involved in the display of the normal pattern of sexual behavior.

**Experiential variables.** Although it has been pointed out that differences in patterns of sexual behavior are associated with differences in genetic background, this does not preclude the importance of experiential factors in determining the particular pattern displayed by the individual. The limits to which experience can modify the behavior pattern are, however, largely determined by genetic factors. In studying the influence of experiential factors on mating behavior, we are indirectly examining the variability allowed by genetic endowment.

When males from two inbred strains of guinea pigs were separated from their mothers at the age of 25 days and reared in isolation, they did not display normal mating behavior patterns when tested as adults. For isolation to produce comparable effects in male guinea pigs of a heterogeneous stock, the animals had to be isolated beginning at 10 days of age. The experiment demonstrates the influence of genetic and experiential variables in determining patterns of sexual behavior. It should be noted that not all males even within the inbred strains showed the "isolation effect" and that the relevant variables associated with isolation have not been determined (Valenstein et al. 1955).

Isolation of the male rhesus monkey at birth produces drastic effects not only on its sexual behavior but on its entire social behavior repertoire (Harlow & Harlow 1962). However, mother-deprived infants, if given an opportunity to interact with peers, develop essentially normal patterns of sexual behavior. The manner in which peer interaction fosters development of normal sexual behavior remains to be demonstrated.

Female guinea pigs and rhesus monkeys reared in isolation also show deficiencies in reproductive behavior. Despite the bizarre social behavior and abnormal sexual behavior displayed by the isolated female rhesus, a number have mated with particularly capable males and have delivered viable young. It should be noted that the behavioral deficiency is not attributable to gonadal hormone deficiency in the isolated monkey or guinea pig.

Although the exact nature and extent of the effects of isolation on sexual behavior have not been determined for the male rat, available evidence suggests that the impairment is not as profound as it is in the monkey (Zimbardo 1958).

A wide variety of experiences are capable of modifying sexual behavior in addition to the method



of rearing just discussed. What has been demonstrated is that sexual behavior, in general, is subject to modification just as other aspects of behavior can be modified by varying experience. Thus despite heritability of sexual behavior patterns, dependence on hormonal stimulation, and the spinal reflex contribution to patterns of sexual behavior, the complete pattern can be blocked, enhanced, or modified by social and situational experience. The extent to which it can be modified by experience varies primarily with the species, sex, and age of the individual.

CHARLES H. PHOENIX

[Other relevant material may be found in GENETICS; INFANCY, article on THE EFFECTS OF EARLY EXPERIENCE; NERVOUS SYSTEM; PSYCHOLOGY, article on COMPARATIVE PSYCHOLOGY; and in the biography of KINSEY.]

#### BIBLIOGRAPHY

- BEACH, FRANK A. 1947a Evolutionary Changes in the Physiological Control of Mating Behavior in Mammals. *Psychological Review* 54:297-315.
- BEACH, FRANK A. 1947b A Review of Physiological and Psychological Studies of Sexual Behavior in Mammals. *Physiological Review* 27:240-307.
- BEACH, FRANK A. 1949 A Cross-species Survey of Mammalian Sexual Behavior. Pages 52-78 in Paul H. Hoch and Joseph Zubin (editors), *Psychosexual Development in Health and Disease*. New York: Grune & Stratton.
- BEACH, FRANK A. 1951 Instinctive Behavior: Reproductive Activities. Pages 387-434 in S. S. Stevens (editor), *Handbook of Experimental Psychology*. New York: Wiley.
- BEACH, FRANK A. 1958 Evolutionary Aspects of Psychoendocrinology. Pages 81-102 in Anne Roe and George Gaylord Simpson (editors), *Behavior and Evolution*. New Haven: Yale Univ. Press.
- BROOKHART, J. M.; and DEY, F. L. 1941 Reduction of Sexual Behavior in Male Guinea Pigs by Hypothalamic Lesions. *American Journal of Physiology* 133:551-554.
- BROOKHART, J. M.; DEY, F. L.; and RANSON, S. W. 1940 Failure of Ovarian Hormones to Cause Mating Reactions in Spayed Guinea Pigs With Hypothalamic Lesions. *Society for Experimental Biology and Medicine, Proceedings* 44:61-64.
- BURNS, R. K. 1961 Role of Hormones in the Differentiation of Sex. Volume 1, pages 78-158 in William C. Young (editor), *Sex and Internal Secretions*. 3d ed. Baltimore: Williams & Wilkins.
- DILGER, WILLIAM C. 1962 Behavior and Genetics. Pages 35-47 in Eugene L. Bliss (editor), *Roots of Behavior: Genetics, Instinct, and Socialization in Animal Behavior*. New York: Harper.
- EATERS, J. T.; and GLASS, A. 1962 The Ovary and Behavior. Volume 2, pages 381-433 in Solly Zuckerman (editor), *The Ovary*. New York: Academic Press.
- FORD, CLELLAN S.; and BEACH, FRANK A. 1951 *Patterns of Sexual Behavior*. New York: Harper.
- GOY, ROBERT W. 1964 Reproductive Behavior in Mammals. Pages 409-441 in Charles W. Lloyd (editor), *Human Reproduction and Sexual Behavior*. Philadelphia: Lea & Febiger.
- GOY, ROBERT W.; and JAKWAY, JACQUELINE S. 1959 The Inheritance of Patterns of Sexual Behaviour in Female Guinea Pigs. *Animal Behaviour* 7:142-149.
- GOY, ROBERT W.; and JAKWAY, JACQUELINE S. 1962 Role of Inheritance in Determination of Sexual Behavior Patterns. Pages 96-112 in Eugene L. Bliss (editor), *Roots of Behavior: Genetics, Instinct, and Socialization in Animal Behavior*. New York: Harper.
- GOY, ROBERT W.; and PHOENIX, CHARLES H. 1963 Hypothalamic Regulation of Female Sexual Behaviour: Establishment of Behavioural Oestrus in Spayed Guinea-pigs Following Hypothalamic Lesions. *Journal of Reproduction and Fertility* 5:23-40.
- HARLOW, HARRY F.; and HARLOW, MARGARET K. 1962 Social Deprivation in Monkeys. *Scientific American* 207, Nov.: 136-146.
- HARRIS, G. W.; MICHAEL, R. P.; and SCOTT, PATRICIA P. 1958 Neurological Site of Action of Stilboestrol in Eliciting Sexual Behaviour. Pages 236-251 in Ciba Foundation, *Symposium on the Neurological Basis of Behaviour*. Edited by G. E. W. Wolstenholme and C. M. O'Connor. Boston: Little.
- JAKWAY, JACQUELINE S. 1959 Inheritance of Patterns of Mating Behaviour in the Male Guinea Pig. *Animal Behaviour* 7:150-162.
- KINSEY, ALFRED C. et al. 1948 *Sexual Behavior in the Human Male*. Philadelphia: Saunders.
- KINSEY, ALFRED C. et al. 1953 *Sexual Behavior in the Human Female*. Philadelphia: Saunders.
- KLÖVER, H.; and BUCY, P. C. 1939 Preliminary Analysis of Functions of Temporal Lobes in Monkeys. *Archives of Neurology and Psychiatry* 42:979-1000.
- LLOYD, CHARLES W. (editor) 1964 *Human Reproduction and Sexual Behavior*. Philadelphia: Lea & Febiger.
- OLDS, JAMES 1958 Effects of Hunger and Male Sex Hormone on Self-stimulation of the Brain. *Journal of Comparative and Physiological Psychology* 51:320-324.
- PHOENIX, CHARLES H. 1961 Hypothalamic Regulation of Sexual Behavior in Male Guinea Pigs. *Journal of Comparative and Physiological Psychology* 54:72-77.
- PHOENIX, CHARLES H. et al. 1959 Organizing Action of Prenatally Administered Testosterone Propionate on the Tissues Mediating Mating Behavior in the Female Guinea Pig. *Endocrinology* 65:369-382.
- SAWYER, CHARLES H. 1960 Reproductive Behavior. Volume 2, pages 1225-1240 in *Handbook of Physiology. Section 1: Neurophysiology*. Washington: American Physiological Society.
- SCHREINER, LEON, and KLING, ARTHUR 1956 Rhinencephalon and Behavior. *American Journal of Physiology* 184:486-490.
- SOULAIRAC, A.; and SOULAIRAC, M.-L. 1956 Effets de lésions hypothalamiques sur le comportement sexuel et le tractus génital du rat male. *Annales d'endocrinologie* 17:731-745.
- VALENSTEIN, ELLIOT S. RISS, WALTER; and YOUNG, WILLIAM C. 1955 Experiential and Genetic Factors in the Organization of Sexual Behavior in Male Guinea Pigs. *Journal of Comparative and Physiological Psychology* 48:307-403.
- YOUNG, WILLIAM C. 1941 Observations and Experiments on Mating Behavior in Female Mammals. *Quarterly Review of Biology* 16:135-156; 311-335.

- YOUNG, WILLIAM C.; GOY, ROBERT W.; and PHOENIX, CHARLES H. 1964 Hormones and Sexual Behavior. *Science* 143:212-218.
- ZIMBARDO, PHILIP G. 1958 The Effects of Early Avoidance Training and Rearing Conditions Upon the Sexual Behavior of the Male Rat. *Journal of Comparative and Physiological Psychology* 51:764-769.

## II SOCIAL ASPECTS

Social scientific writing on sexual behavior over the past century falls in three categories: the ecology and styles of sex as a physical act; the meanings of the act, in its direct, displaced, or sublimated form, to the individual and to society; the norms and ethical standards that govern the modes of, occasions for, and designation of participants in, the physical act; and the organizational arrangements that implement these norms and standards. It is now commonly accepted that the frequencies and techniques of sexual behavior vary widely among individuals and among religious, class, ethnic, educational, age, and sex categories in the same society (Kinsey et al. 1948; 1953); that its personal meanings are numerous and contradictory (Freud 1893-1895); and that sexual norms and ethical standards contrast sharply from culture to culture (Mantegazza 1886).

As producers of literature on sexual behavior, social scientists are easily outnumbered by physicians, psychologists, physiologists, theologians, and philosophers, among others. The present article, rather than restricting itself to studies by social scientists, will conceptualize the content of this diverse literature from a social science perspective. These materials will be examined in terms of the social meanings of sexual behavior, the way the sexual enters as a component of social roles, the social norms governing recruitment to these eroticized roles, and their articulation with networks of the noneroticized roles that constitute some major social institutions. Reference will be made to social organizational arrangements for facilitating or inhibiting enactment of these roles.

**The social science perspective.** Sexual behavior participates in social action by contributing a motive, or a driving force of that action. Ratzenhofer (1898), an early theorist of the sociology of sex, saw all motives of social contact as modifications of either the instinct of self-preservation or the sexual instinct. Moreover, as a component of social action, sexual behavior is also a way of expressing and conveying both individual and social meanings; that is, the meaning is not only that which an individual attributes to his act but

also that which society has come to apply to the act (Thomas 1907).

In contrast to the social science perspective, the "physicalist tradition" tends to treat sex simply as energy or motive. For example, the studies of Kinsey (a zoologist) and his associates at the Indiana University Institute for Sex Research follow the tradition of "biological realism"; that is, they focus on the ecology and styles of the physical act. These are by far the most popular scientific studies of sex, as measured by sales and published critical reviews. From a physicalist point of view, the presumption that all acts engaging the genitals are alternative ways of tapping the same reservoir of physiological energy justifies drawing them under the same rubric. Thus Kinsey's treatment of alternative "outlets" places greater emphasis on the channels through which energy is released than on the social context of that energy release.

Social science analyses of sexual behavior are primarily concerned with coital acts, both homosexual and heterosexual, which involve a relationship between human beings and are thus qualitatively different from other sexual behavior. Masturbation, nocturnal emissions, and bestiality claim social science attention only insofar as they affect either the occurrence of coitus or its meaning or become the object of institutional concern. According to this view, the energy problem is interwoven with that of the social meaning of sexual acts. Freudian theory of the economy of the libido uses the notion of sublimation to designate the rechanneling of sexual energy into cultural activities by shifting the object of cathexis; thus the theory interweaves motivational and meaningful aspects.

Sexual behavior contains a nonverbal language through which an individual expresses and conveys meanings; sensory exploration and gestures outweigh words as sign and symbol vehicles. Removal of clothing beyond the degree necessary for coitus itself exemplifies the significance of tactile communication. Any sexual act may carry several social and personal meanings. For example, a morning erection may express unconscious concerns about death rather than sex (Bergler 1950).

Ford and Beach (1951) studied sexual behavior between species and across cultures, yet they granted priority to the physicalist approach by classifying sexual behavior according to modes of stimulation of the genitals. A consequence of this physical perspective is their tendency to measure the cause of sexual behavior in terms of tension reduction and to designate procreation as its prin-



cipal social function. The procreative function loses its centrality for a social psychology of human sexual behavior in view of the empirical fact that less than one in a thousand human coital acts results in pregnancy and fewer are intended to do so (Foote 1954). However, for sociology, the part of sexual behavior that is, in fact, procreative becomes significant because of both the demographic consequences and the institutional and organizational arrangements that become necessary for the birth and rearing of offspring (Ellwood 1925).

**The eroticized role.** The higher a species is in the evolutionary scale, the less dependent is its sexual behavior on gonadal hormones and seasonal physiological changes; and in man it comes under control of the cortex of the brain (Ford & Beach 1951), thus introducing the possibility of decisional and normative control. Sexual behavior as a role component, that is, as a component of social action under decisional and normative control, distinguishes human from animal sexuality.

In human sexual behavior, the erotic element may be incidental to other motives, such as the desire for a new experience or the desire to escape drudgery (Thomas & Znaniecki [1918-1920] 1958, vol. 2, pp. 1800-1821). A social role—such as that of lover, of prostitute or client, of husband or wife, of young male delinquent or adult fellator (A. J. Reiss 1961)—in which a sexual element is embedded with other social role elements, such as economic or affective exchange or child raising, may be termed an "eroticized role." The eroticized role is part of the whole system of roles constituting the personality, and, in a sense, it eroticizes the entire personality. This is what has been described as a transformation of sexuality into Eros (Marcuse 1955). Sexuality becomes meaningful to the actor, and may be understood by the observer, in the context of these other aspects of the personality.

An eroticized social role may be articulated with other roles in the social structure; for example, the role of the prostitute in her relation with a client may be meshed with her relatively non-erotic role in her relation with her madam or other employer. This perspective on role articulation follows Malinowski (1929), who advocated an anthropological perspective that does not treat sex as a mere physiological transaction but studies its implication for lovemaking, its function as the nucleus of the family, the spells and magic that grow up around it, its effects upon, and how it is affected by, the legal system and the economic system.

The variety of sexual meanings is, in part, dependent on other aspects of the role—that is, on other norms, affects, or meanings that define that role or other roles with which it is articulated in a personality or in a society. Kirkendall (1961) described how sex, while it may advance a deep interpersonal relation, may be exploitative as part of a casual relationship. With prostitutes, the relationship itself becomes the instrumental servant of its sexual component. Even in advancing a deep relation, however, sex may assume special meanings associated with the broader meaning of that relationship. Thus sex may be a weapon to destroy, a mode of defense, a bargaining point, or a form of self-assertion; negatively, it may be a form of self-denial (Frank 1954).

Components of an eroticized role may be malintegrated. A marriage contracted for social or economic advantages despite sexual incompatibility is a classic instance. Sexual behavior inconsistent with personal ethical standards may engender an internal struggle manifested in guilt feelings. If strains of this type become widespread, social standards may be readjusted (I. L. Reiss 1960).

### Recruitment to the eroticized role

Recruitment of partners to an eroticized role is governed by considerations of personality, social norms, and societal mechanisms of control. Personality variables affect the desire and ability of an individual to seek, as well as his desirability as, a sex-role partner. Physiological or attitudinal disturbances in an individual's response at any stage of the coital act influence recruitment to the eroticized role. In the preparatory stage, an individual must be able to respond to excitation from internal and external psychological stimulation. This requires sensory, intellectual, motor, and glandular contributions to mechanisms of arousal. An individual's attitude toward himself—for example, as reflected in body image—affects his ability to woo another through foreplay and enter intimate relationships (Fisher & Cleveland 1958). An individual must experience impetus to penetrate or desire to be penetrated, want to continue inplay, and, at culmination, experience orgasmic peristalsis of genital structures. Much depends upon ability to tolerate regression in the service of the ego (Fried 1960). Disturbances of sexuality are commonly classified in terms of frigidity, low potency, preference for masturbation or for sexual perversion, hypersexuality, and lack of sexual interest (Eisenstein 1956).

Special personality groups vary in interest in eroticized roles. Most of the erotic impulse of idiots

is directed toward themselves and the remainder to objects of the same sex (Potter 1927). Alcoholics lack sex interest, because, it has been argued, of repressed homosexuality (Levine 1955). The aged tend to lose motivation for coitus (Armstrong 1963).

The evaluation of personal characteristics in recruitment is relative to the persons involved; that is, an individual not desirable to one partner may be desirable to another. For example, males generally prefer shorter females, and females generally prefer taller males. Women who want men near their own height, it has been argued, are expressing a need for ascendancy. Similarly, males who seek taller women are said to be seeking to possess the forbidden parent figure (Beigel 1954).

**Norms of recruitment.** Social norms prescribe and proscribe which social types may recruit one another. Havelock Ellis (1910), for example, wrote that in choosing a mate we tend to seek parity of racial and anthropological characteristics together with disparity of secondary sexual characteristics and complementarity of psychic characteristics. Under control of instinct, animals rarely depart from parity of species and disparity of secondary sexual characteristics. Mongrelization may be induced among domesticated animals but is extremely rare among wild animals despite opportunities. Man can, and sometimes does, engage in relations that the more constitutionally directed animals would refuse. Man may disregard species parity and practice bestiality, disregard secondary sexual disparity and practice homosexuality, and disregard psychic complementarity and establish neurotic sex relations. On the other hand, man may refuse relations because of sexual style, the status of the partner, or the beauty of the partner—a parity irrelevant to the animal. For man, the power of the social over biological determinants extends to the very definition of gender. Money et al. (1957) compared the relative influences of the chromosomal sex, gonadal sex, hormonal sex, internal accessory reproductive structures, and external genital morphology on the gender role actually assumed by hermaphrodites, and they found social sex assignment to be the most reliable prognosticator of gender role.

Rules for recruitment to coital relations in general parallel, but do not completely overlap with, those governing marital recruitment. Laws of exogamy define the inner social limits for recruitment of a marital partner. These are extensions of the incest taboo, which is a nearly universal taboo, with exceptions only in such cases as the Egyptian royal family, where brother-sister mar-

riage was permitted, or among Azande nobles, where father-daughter marriage has been permitted. Laws of endogamy define the outer social limits. These norms have been compared by Westermarck (1889) to the "law of similarity" that keeps animals from pairing across species. In human relations, however, endogamy controls potentially fertile relations within the same species.

A number of individuals are virtually precluded from recruitment to eroticized heterosexual roles. This is the case with institutionalized populations in prisons or mental hospitals. Likewise, occupational conditions may separate individuals from the opposite sex. Hobos, being homeless, have less access to women, and thus they tend to frequent the lowest status prostitutes and to practice homosexuality.

Social control of eroticized role recruitment contributes to the maintenance of the social structure; if society lacked such controls, random mating could disrupt the familial and economic lines of stratification of a society. Moreover, social control of recruitment may enable one social group to dominate another through its power to allocate sexual occasions and facilities. It may also be an internally integrative device, as it is among celibate monastics. According to Ross (1920), such control is necessary to manage a smoldering antagonism that exists between the sexes as groups.

**Sexual communication in recruitment.** Communication employed for recruitment to the eroticized role parallels the communication of arousal in the sex act. Token arousal announces sexual availability. Erotic communication relies heavily on movement, such as the inviting movements of the dance and the arousing tactile movements during coitus. Odors founded in the natural capryl odors and supplemented at the human level by perfumes act upon the deeper levels of consciousness (Bloch 1907). Cosmetic means of communicating sexual interest include painting and tattooing as well as sexual mutilation. Socially stylized ceremonies of fertility or initiation may announce availability and effect arousal. A few rites, such as subincision, announce sexual availability but reduce sexual enjoyment or arousal potential for the subject (Allen 1949; Westermarck 1889).

Verbal communication supplements sensory communication in sexual behavior. Speech is so closely identified with sexuality that in some cultures it is forbidden or restricted between persons who are socially precluded from engaging in sexual relations (Baker 1949). The more intimate a sexual matter, the less the likelihood of communication about it to nonerotic role partners.



In many societies, socially structured mores about erotic communication bar young people from arousal and consummatory opportunities. Sexual norms may be learned inferentially by the young or by new social groups, such as immigrants, by observation of behaviors lacking primary erotic meaning, such as forms of dress and of association between the sexes (Klausner 1964). Classified newspaper advertising and a specialized journalism directed to erotica have long been used for the recruitment of both heterosexual and homosexual partners.

Deviant subgroups develop private signaling systems to cue a potential partner without revealing themselves to noninitiate spectators. For example, homosexuals may wear peculiar clothing (the definition of which changes from time to time), intersperse their speech with a special argot, or perform certain gestures. They may position themselves to receive such signals by frequenting known haunts.

### Social control of sexual behavior

Society not only exerts control over the recruitment of role partners but also attempts to control the types of sexual behaviors enacted and the occasions for their enactment. Both informal and formal social controls are found. Attitudinal disapproval of individuals' erotic behaviors, which is perhaps the commonest form of control, may be expressed as verbal chastisement, gossip, scandal, humor, or mockery. The Trobriand Islanders, for example, control sexual style by jesting about those who use the "white man's position" for intercourse. A quasi-formal control mechanism is exemplified among the Mondurucu by the gang rape to force a recalcitrant female to submit to male authority (Murphy 1959).

Norms also define those behaviors to which it is legitimate to recruit. Early studies of sexual behavior treated cross-cultural comparisons descriptively, presenting various behaviors, sexual art, and technical devices as curiosities (Moll 1912; Kraftt-Ebing 1886; Bloch 1907). These studies did, however, open the way to the study of sexual norms by giving information about criteria for recruitment to eroticized roles in other cultures. Correlations between sexual behavior and the overarching cultural ethos allow inference about the types of sexual behaviors to which recruitment is approved in various cultural situations; that is, recruitment norms vary with the ethos. For example, Sorokin (1956) argued that sexual asceticism is associated with "idealistic societies" and libertarianism with "sensate societies." Taylor (1953) classified histor-

ical periods in Western society as predominantly "matristic" or "patristic." In patristic periods, dominated by a father religion and exemplified by England during the first millennium of the Christian era, people have a horror of homosexuality, sex is generally restricted, and rape and incest increase. In matristic periods, dominated by a mother religion and exemplified by England after the first millennium, people have a horror of incest and are permissive regarding sex.

Social control may be exerted by concrete organizational arrangements. Organizations that facilitate recruitment to the sex role also maintain procedures, such as the "sex game," for controlling recruitment and defining types of permitted alliances (e.g., see Cressey 1932). The Nazis, while promoting a policy of sexual abstinence, provided youth camps in which sexual relations took place. Monasteries may debar their inmates from sexual opportunities. Child marriage commits the relationship before the erotic drive appears. The sexes may be physically and, therefore, socially isolated, as in a harem or school. Chaperonage is a way of supervising the young during their meetings. Systems of peer group pressures, such as those exerted by campus sororities or fraternities, encourage socially approved love relationships by subjecting them to ratings by age mates. These controls are more stringent among the upper classes, who have more to lose by a breakdown in the stratification system (Goode 1959).

Legal control takes the form of laws against homosexuality, adultery, sodomy, or pederasty—that is, against behaviors that violate rules governing recruitment to the eroticized role. In some jurisdictions, the style of coitus is legally as well as informally directed. In England, the Wolfenden Committee on homosexual offenses and prostitution gathered intelligence about sexual behavior from social scientists and psychiatrists as a basis for legislating on sexual recruitment. They concluded that it is the duty of the law not to concern itself with immorality as such but to confine itself to activities that offend against public order and decency. Thus, they recommended that homosexual relationships between adults by mutual consent and in privacy should not fall within the province of criminal law (Great Britain . . . 1957). Federal legislation in the United States, such as the Mann Act (the so-called "white slave" traffic act) of 1910, is given force through policing activities of the Immigration and Naturalization Service, the Federal Bureau of Investigation, and local police departments. Throughout Anglo-American history,

the methods of legal control of sex expression have varied much more than has the accepted doctrine of sexual morality (May 1931; Mueller 1961).

Sex education is a form of noncoercive, though formal, control. The medical profession in its efforts to control venereal disease, as well as religious groups in their efforts to maintain their own sexual norms, have been instrumental in founding organizations such as the American Society of Sanitary and Moral Prophylaxis in 1905 and the American Social Hygiene Association in 1912, now the American Social Health Association. This latter association uses education to combat venereal disease, commercialized prostitution, and other conditions associated with sex delinquency among young people.

**Resistance to social control.** Protest against efforts to enforce prevailing sexual mores and to control recruitment to the eroticized role is expressed in individual sexual deviance, the formation of groups of individuals with special sexual needs, and movements to change the prevailing mores. The sex offender uses illegitimate means, such as violence or disregard for the consent of the partner, to recruit to coital or noncoital sexual behaviors, or he recruits without regard to rules defining permitted sexual partners. Generally, these individuals are undersexed rather than oversexed. Freudian theory suggests that some sex offenders may be compensating for feelings of bodily damage or phallic inadequacy, or suffering from castration anxiety (Hammer 1957). Most sex offenders are of average intelligence. Those arrested for incest, however, tend to be older men of subnormal intelligence and antisocial personalities. Apprehended distributors of obscene literature tend to be older, sexually inhibited men of average intelligence who are not hostile. Women are rarely arrested for sexual offenses in American society, except for violating norms regarding public solicitation. The role relations in primarily nonsexual delinquent groups may become eroticized, involving, for instance, group masturbation.

Organizations of individuals whose physical status makes it difficult for them to recruit or be recruited to an eroticized role, such as dwarfs or the blind, may oppose prevailing sexual mores, as may those who, though not necessarily handicapped or abnormal in a clinical sense, have sexual requirements that are contrary to the norms. These groups provide evaluational support, facilitate recruitment to the role, and insulate members against the sexual mores of the larger society (Hooker 1956).

Social movements concerned with sex have sought revision of sexual standards, of the laws supporting them, and of the means of enforcing them. Corruption in systems of police regulation of prostitution led to the formation of the International Federation for the Abolition of State Regulation of Vice in 1875. The World League for Sexual Reform on a Scientific Basis, perhaps the most extensive such group, was an outgrowth of the Institute for Sexual Science founded by Magnus Hirschfeld in Germany in 1929. It attempted to abolish antisexual taboos, establish a philosophical basis for sexual freedom, and provide education in the form of lectures and clinical advice. The institute was closed by the Nazis, and the world movement collapsed a few years later. Wilhelm Reich, who criticized Hirschfeld for attempting sexual reform within the framework of capitalist society, established the Socialist Society for Sexual Advice and Study in Vienna in 1928 to "defend the rights of children and young people to their natural love."

### Sex and social institutions

The nonerotic components in an eroticized role may link it with a pattern of roles constituting an institution. Historically, eroticized roles have been part of the fabric of economic, political, religious, familial, and other institutions. At certain times, a particular institution, such as the church in Christian society, may exclude the erotic component from its roles and influence its place in the roles of another institution, especially the family.

**The family.** Social movements to reform sexual mores ordinarily protest the family's monopoly of the legitimate eroticized role. The eroticized role is universally legitimated in the family, and in modern Western society it is legitimated only in the family. However, actual behavior is at variance with this official standard. In the Kinsey samples, 50 per cent of the females and 85 per cent of the males with a high school education reported that they had experienced premarital coitus. Nearly half of these females had experienced coitus with a man other than their fiancé. Moreover, by the age of forty nearly 26 per cent of the females and 50 per cent of the males had engaged in extramarital coitus.

Increasing opportunities for premarital and extramarital coitus accompany the greater mixing of the sexes in the educational, occupational, and social worlds. Legal control of premarital intercourse in American society is generally enforced more strictly in the case of younger participants.



Adultery is grounds for divorce in all states. To the extent that coitus takes place outside of marriage the procreative aspect declines in significance. For example, Malinowski (1929) thought that the Trobriand Islanders, who have very permissive sexual mores, did not appreciate the connection between coitus and pregnancy, although it is more likely that, while aware of the connection, they were not much interested in it. Discussions of contraception reveal some tensions aroused concerning nonprocreative coital meanings, since contraceptives are not simply designed to prevent conception but to do so while coitus continues.

**Sex and religion.** Sexual behavior has been interwoven with religion from earliest times. Religion's general concern with fertility involves it in the procreative aspect of sex, and this concern may even become pervasive, as, for example, in the Indian *lingam* cults. Furthermore, sexual energy itself may play a role in religious worship. Scott (1941) concluded that phallic worship originated in the pleasure associated with coitus and not in any clearly conceived notion that intercourse would produce children. Sex as worshipful, expressive communication with a supraindividual force is illustrated by the activities of temple prostitutes in Babylon. Krafft-Ebing (1886) argued that religion and sex could replace one another, since both display a similar scope and quality of excitement at their peaks and both can degenerate under pathological conditions into cruelty. Religious mystical roles may contain an erotic compound in both the Western and Eastern traditions. On the other hand, the cases of the sacerdotal celibate, the religious hermit, or the penitent abstainer illustrate religious exclusion of the sexual component.

Religious considerations affect recruitment to eroticized roles in other institutions. Thus, rules of exogamy and endogamy are generally supported by religious sanctions. Religious sanctification of marriage or the withholding of that sanctification affects the recruitment of coital partners. Moreover, religion may contribute to the isolation of the sexes by separating them in worship and by excluding women from sacred precincts.

Religious devotion is negatively associated with participation in nonreligious eroticized roles in the United States. The frequencies for all sexual outlets, except for marital coitus, decrease with increasing devoutness (Kinsey et al. 1948; 1953). This association seems to hold in France as well (Institut Français . . . 1960).

**Sex and the polity.** The relation of sexual behavior to the nature of the state and politics has long been of popular concern. The literature on

court ladies and on espionage has described the impact of eroticized roles upon political roles. Sorokin (1956) asserted that sexual overindulgence debilitates the entire society. Krafft-Ebing (1886) saw a two-directional relation, in which, on the one hand, sexual excess undermines the props of society and, on the other hand, the collapse of society produces sexual aberrations.

Utopian schemes have included recommendations for control of eroticized roles. Unwin (1940) proposed that, in his new society, Hopousia, there be two kinds of marriage: alpha marriage, which would be monogamous and preceded by prenuptial continence, for those who would achieve, create, and lead, and beta marriage, which may be terminated at will and would not require prenuptial continence, for those who do not aspire to social position. Unwin's thesis is grounded in the concept of energy sublimation. In John Humphrey Noyes' Oneida community, a cooperative colony founded in 1848, extramarital "coitus interruptus" was permitted, on the ground that sex is a way of communicating the meaning of intimate friendship.

The acquisition of political power by men usually implies an access to more women. For example, in Bedouin society, a tribal chief is more likely to be polygamous and to retain concubines than is a less powerful member of the tribe. In Europe, royalty as well as the economically powerful have been able to keep mistresses. Conquest has generally implied access to the women of the conquered society, either by rape, by capture, or through the woman's desire to consort with the powerful.

**Sex and the economy.** Sex has a place in the economy both as a marketable commodity and as a component affecting the marketability of nonsexual commodities. Thus economic value can be derived from sexual value. For example, prostitution involves an economic-sexual exchange of this character; the economic exchange ramifies beyond the prostitute and her client to include the pimp who recruits the client, the madam who supplies facilities for the act, and, perhaps, the police who protect all of them from the law. In the case of marriage, the economic-sexual exchange is embedded in a wider set of family and community relations. Sexual role performance consumes facilities, goods, and services, and thus industries emerge around the provision of facilities for recruitment to the eroticized role and for enactment of that role. These range from the provision of meeting places, such as dance halls; the publishing of pornographic, scientific, artistic, and guidance manuals; the supply of pharmaceutical and me-

chanical contraceptives; and the letting of places, such as a motel room or brothel, for consummating the sex act.

Sexual value may be derivative from economic value as well. For example, sexual jealousy, rather than being an innate emotion, may be related to views of women as property. Jealousy is then an anger aroused when property rights are violated. A norm of premarital chastity may also be related to marketability (Nemeček 1958).

Reich (1930), using syncretistic Marxian-Freudian "sex-economic" concepts, described an interplay between sexual and economic norms. He reasoned that since psychic structure is created by social structure and the core of psychic structure is the sexual function, sexual function is controlled by social structure. Social structure is constituted by the "modes of production," including the interpersonal relations of production. These relations affect sexual needs, among others. They also influence concepts of life, morals, and philosophy—that is, ideology—and so influence sex norms. Classes standing in different relation to the "modes of production" will differ in their sexual needs and norms. The dominant political minority imposes its ideology in general, and its sexual ideology in particular, on a population. According to this view, the subordinate population thus has sex needs derived from its own class position while being subject to the sexual ideology of another class. It should be noted that Kinsey's findings do not support Reich; Kinsey and others (1953) found little relation between the parental occupational class and the incidence of various sexual behaviors.

### Research methods

Methods of social and psychological research developed in other areas have been used to ascertain frequencies of sexual behaviors as well as social and individual meanings associated with that behavior. Questionnaires (Christensen & Carpenter 1962) and interviews (Kinsey et al. 1948) have both been used to elicit such information. Protecting the anonymity of the respondent and establishing the scientific legitimacy of the interviewer are salient bases for rapport during an interview. To maximize the validity of responses, the interviewer must avoid making the respondent feel guilt, anxiety, or defensiveness. Direct questions should be asked with no defensiveness or apology on the part of the interviewer, and terminology should be adapted to the social-class level of the respondent. The inquiry should begin with matters that do not provoke anxiety, such as age, and then proceed to items which are sexual but for

which the individual does not feel responsible; thus the interviewer would ask about spontaneous orgasms before asking about masturbation. The fact that anxiety associated with each type of behavior differs among culture or class groups and between the sexes affects the ordering of the questions. For example, homosexuality is less frightening than masturbation for females in American society, while the reverse is true for males; thus, in interviewing females, questions on homosexuality precede those on masturbation, and the reverse order is used for males. As a practical procedure one might request the age at first orgasm, then ask for its source and, if this is not too anxiety-provoking, trace this source through the years. By asking when he first experienced each type of behavior, the burden of denial is placed upon the interviewee. Kinsey's interviewers severely censured any subject suspected of giving a dishonest history.

Kinsey correlated the frequencies of each source of outlet—masturbation; premarital, extramarital, and marital coitus; homosexuality; and animal contacts—with demographic factors such as national origin, marital status, age, education, occupational class, rural or urban residence, religious affiliation, and religious devoutness, among others. Intrapsychic meanings of sexual behavior have been studied by correlating projective test responses with the frequencies of behaviors. The Rorschach test has been used to reveal body image (Fisher & Cleveland 1958), the Thematic Apperception Test to uncover feelings of guilt associated with sex (Leiman & Epstein 1961), and a combination of the Thematic Apperception Test and the House-Tree-Person Test to obtain data about castration feelings or feelings of phallic inadequacy for psychoanalytic interpretation (Hammer 1957). Anthropological field methods, employing observation and informant interviews, have been used to gain information about both the occurrence of behaviors and their meanings (Malinowski 1929). Observational methods supply data on systems of communication and interaction that are necessary for an analysis of the social organization of sexuality (A. J. Reiss 1961). The Human Relations Area Files, containing field reports on many societies, have been used to exploit the advantages of both field observation and correlation methods (Murdock 1949). Attitude questionnaires have been used to obtain information on ideal prescriptive norms, as has content analysis of newspapers, radio, and plays (A. Ellis 1954).

The fact that the various sexual behaviors may be ordered along dimensions of greater or lesser intimacy or greater or lesser acceptability has been



exploited to construct scales. Podell and Perkins (1957), for example, developed a Guttman scale for ordering types of sexual experience, and Ira Reiss (1964) developed a refined Guttman scale for a study of standards of sexual permissiveness.

SAMUEL Z. KLAUSNER

[See also FAMILY; MARRIAGE; PROSTITUTION; SEXUAL BEHAVIOR, articles on SEXUAL DEVIATION and on HOMOSEXUALITY; and the biographies of ELLIS; FREUD; MALINOWSKI; REICH; WESTERMARCK.]

#### BIBLIOGRAPHY

Primary literary materials as sources of data may be located in several extensive bibliographies: Hayn & Gutendorf 1875 lists references and abstracts of erotica in German; Reade 1936 lists and abstracts English material; and Gay 1861 includes a similar collection of French materials. Ellis & Abarbanel 1961, containing articles and short bibliographies on most topics of interest, is a good introductory resource for secondary materials.

- ALLEN, CLIFFORD 1949 A Study of Sexual Mutilations. *International Journal of Sexology* 2:158-161.
- ARMSTRONG, EUNICE B. 1963 The Possibility of Sexual Happiness in Old Age. Pages 131-137 in *Society for the Scientific Study of Sex, Advances in Sex Research*. Edited by Hugo G. Beigel. New York: Harper.
- BAKER, SIDNEY J. 1949 Speech and Sexual Taboos. *International Journal of Sexology* 3:98-101.
- BEIGEL, HUGO G. 1954 Body Height in Mate Selection. *Journal of Social Psychology* 39:257-268.
- BERGLER, EDMUND 1950 Morning Erections. *International Journal of Sexology* 3:141-144.
- BLOCH, IWAN (1907) 1937 *The Sexual Life of Our Time*. New York: Falstaff. → First published as *Das Sexualleben unserer Zeit in seinen Beziehungen zur modernen Kultur*.
- CHRISTENSEN, HAROLD T.; and CARPENTER, GEORGE R. 1962 Timing Patterns in the Development of Sexual Intimacy: An Attitudinal Report on Three Modern Western Societies. *Marriage and Family Living* 24:30-35.
- CRESSEY, PAUL G. 1932 *The Taxi Dance Hall*. Univ. of Chicago Press.
- EISENSTEIN, VICTOR W. (editor) 1956 *Neurotic Interaction in Marriage*. New York: Basic Books.
- ELLIS, ALBERT (1954) 1962 *The American Sexual Tragedy*. 2d ed. New York: Stuart.
- ELLIS, ALBERT; and ABARBANEL, ALBERT (editors) 1961 *The Encyclopedia of Sexual Behavior*. 2 vols. New York: Hawthorn.
- ELLIS, HAVELOCK (1910) 1936 *Studies in the Psychology of Sex*. Volume 4: Sex in Relation to Society. New York: Random House.
- ELLWOOD, CHARLES A. 1925 *The Psychology of Human Society*. New York: Appleton.
- FISHER, SEYMOUR; and CLEVELAND, SIDNEY E. 1958 Body Image Boundaries and Sexual Behavior. *Journal of Psychology* 45:207-211.
- FOOTE, NELSON N. 1954 Sex as Play. *Social Problems* 1:159-163.
- FORD, CLELLAN S.; and BEACH, FRANK A. 1951 *Patterns of Sexual Behavior*. New York: Harper.
- FRANK, LAWRENCE K. 1954 The Psychocultural Approach in Sex Research. *Social Problems* 1:133-139.
- FREUD, SIGMUND (1893-1895) 1955 *The Standard Edition of the Complete Psychological Works of Sigmund Freud*. Volume 2: Studies on Hysteria, by Josef Breuer and Sigmund Freud. London: Hogarth. → First published in German.
- FRIED, EDITHA 1960 *The Ego in Love and Sexuality*. New York: Grune.
- GAY, JULES (1861) 1894-1900 *Bibliographie des ouvrages relatifs à l'amour, aux femmes, au mariage...* 4th ed. 4 vols. Paris: Becour.
- GOODE, WILLIAM J. 1959 The Theoretical Importance of Love. *American Sociological Review* 24:38-47.
- GREAT BRITAIN, COMMITTEE ON HOMOSEXUAL OFFENSES AND PROSTITUTION (1957) 1963 *The Wolfenden Report: Report of the Committee on Homosexual Offenses and Prostitution*. New York: Stein & Day.
- HAMMER, EMMANUEL F. 1957 A Psychoanalytic Hypothesis Concerning Sex Offenders. *Journal of Clinical and Experimental Psychopathology* 18:177-184.
- HAYN, HUGO; and GUTENDORF, ALFRED N. (1875) 1912-1929 *Bibliotheca germanorum erotica et curiosa*. 9 vols., 3d ed. Munich: Muller. → Hayn was the sole author of the first edition, published as *Bibliotheca germanorum erotica*.
- HOOKE, EVELYN 1956 A Preliminary Analysis of Group Behavior of Homosexuals. *Journal of Psychology* 42:217-225.
- INSTITUT FRANÇAIS D'OPINION PUBLIQUE (1960) 1961 *Patterns of Sex and Love: A Study of the French Woman and Her Morals*. New York: Crown. → First published in French.
- KINSEY, ALFRED C. et al. 1948 *Sexual Behavior in the Human Male*. Philadelphia: Saunders.
- KINSEY, ALFRED C. et al. 1953 *Sexual Behavior in the Human Female*. Philadelphia: Saunders.
- KIRKENDALL, LESTER A. 1961 *Premarital Intercourse and Interpersonal Relationships*. New York: Julian.
- KLAUSNER, SAMUEL Z. 1964 Inferential Visibility and Sex Norms in the Middle East. *Journal of Social Psychology* 63:1-29.
- KRAFFT-EBING, RICHARD F. VON (1886) 1965 *Psychopathia sexualis*. New York: Putnam. → First unpurgated edition in English; original in Latin.
- LEIMAN, ALAN H.; and EPSTEIN, SEYMOUR 1961 The-matic Sexual Responses as Related to Sexual Drive and Guilt. *Journal of Abnormal and Social Psychology* 63:169-175.
- LEVINE, JACOB 1955 The Sexual Adjustment of Alcoholics: A Clinical Study of a Selected Sample. *Quarterly Journal of Studies on Alcohol* 16:675-680.
- MALINOWSKI, BRONISLAW (1929) 1962 *The Sexual Life of Savages in North-western Melanesia: An Ethnographic Account of Courtship, Marriage, and Family Life Among the Natives of the Trobriand Islands, British New Guinea*. New York: Harcourt.
- MANTEGAZZA, PAOLO (1886) 1935 *The Sexual Relations of Mankind*. New York: Eugenics Publishing. → First published in Italian.
- MARCUSE, HERBERT 1955 *Eros and Civilization: A Philosophical Inquiry Into Freud*. Boston: Beacon.
- MAY, GEOFFREY 1931 *Social Control of Sex Expression*. New York: Morrow.
- MEAD, MARGARET 1949 *Male and Female: A Study of the Sexes in a Changing World*. London: Gollancz; New York: Morrow.
- MOLL, ALBERT (1912) 1921 *Handbuch der Sexualwissenschaften*. 2d ed. Leipzig: Vogel.

- MONEY, JOHN; HAMPSON, J. G.; and HAMPSON, J. L. 1957 Imprinting and the Establishment of Gender Role. *A.M.A. Archives of Neurology and Psychiatry* 77:333-336.
- MUELLER, GERHARD 1961 *Legal Regulation of Sexual Conduct*. Legal Almanac Series, No. 9. Dobbs Ferry, N.Y.: Oceana.
- MURDOCK, GEORGE P. 1949 *The Social Regulation of Sexual Behavior*. Pages 256-266 in Paul Hoch and Joseph Zubin (editors), *Psychosexual Development in Health and Disease*. New York: Grune.
- MURPHY, ROBERT F. 1959 Social Structure and Sex Antagonism. *Southwestern Journal of Anthropology* 15: 89-98.
- NEMEČEK, OTTOKAR 1958 *Virginity: Prenuptial Rites and Rituals*. New York: Philosophical Library.
- PODELL, LAWRENCE; and PERKINS, JOHN C. 1957 A Guttman Scale for Sexual Experience: A Methodological Note. *Journal of Abnormal and Social Psychology* 54:420-422.
- POTTER, HOWARD W. 1927 An Introductory Study of the Erotic Behavior of Idiots. *Journal of Nervous and Mental Disease* 65:497-507.
- RATZENHOFER, GUSTAV 1898 *Die sociologische Erkenntnis*. Leipzig: Brockhaus.
- READE, ROLF S. 1936 *Registrum librorum eroticorum* . . . 2 vols. London: Privately printed.
- REICH, WILHELM (1930) 1945 *The Sexual Revolution*. 3d ed. New York: Orgone Institute Press. → First published as *Die Sexualität im Kulturkampf*.
- REISS, ALBERT J. Jr. 1961 The Social Integration of Queers and Peers. *Social Problems* 9:102-120.
- REISS, IRA L. 1960 *Premarital Sexual Standards in America*. Glencoe, Ill.: Free Press.
- REISS, IRA L. 1964 The Scaling of Premarital Sexual Permissiveness. *Journal of Marriage and the Family* 26:188-198.
- ROSS, EDWARD A. (1920) 1938 *Principles of Sociology*. 3d ed. New York: Century.
- SCOTT, GEORGE R. (1941) 1952 *Phallic Worship*. New York: Mental Health Press.
- SOROKIN, PITIRIM A. 1956 *The American Sex Revolution*. Boston: Sargent.
- TAYLOR, GORDON R. 1953 *Sex in History*. London: Thames & Hudson.
- THOMAS, W. I. 1907 *Sex and Society: Studies in the Social Psychology of Sex*. Univ. of Chicago Press.
- THOMAS, W. I.; and ZNANIECKI, FLORIAN (1918-1920) 1958 *The Polish Peasant in Europe and America*. 2 vols., 2d ed. New York: Dover. → See especially Volume 2, pages 1800-1821, "Sexual Immorality of Girls."
- UNWIN, JOSEPH D. 1940 *Hoplusia: Or, the Sexual and Economic Foundations of a New Society*. London: Allen & Unwin.
- WESTERMARCK, EDWARD A. (1889) 1921 *The History of Human Marriage*. 3 vols., 5th ed. London: Macmillan.

### III

#### SEXUAL DEVIATION: PSYCHOLOGICAL ASPECTS

One premise of psychoanalytic doctrine is that the general developmental process of childhood includes psychosexual development. It may be taken as a general premise of all growth, physical and psychological, that at any one stage of devel-

opment the organism's current growth status and future potential is determined by (1) what it began with, (2) what has been added or subtracted in the interim, perhaps irreversibly, and (3) the growth-facilitating or growth-impeding environment that it may meet. Thus, at the moment of conception, growth status and potential is determined by both the quantitative and qualitative aspects of the genetic code of information, supplied half by the mother, half by the father, and by the intrauterine environment in which the fertilized egg exists. The intrauterine environment may prove hostile mechanically, metabolically, or because of the invasion of foreign organisms, viruses for example. The embryo itself may be its own enemy, as when a particular developmental phase or system fails, with subsequent deleterious effects on another phase or system. For example, if the testes differentiate imperfectly in a genetic male, then the fetal testicular secretions fail, with the result that the anlagen of the external genital morphology develop as either completely or partially female, depending on the degree of testicular failure. Contrarily, in the case of a genetic female, if the fetal adrenal cortex functions erroneously, so as to produce an excess of androgen in the place of expected cortisone, then the anlagen of the external genitalia will masculinize in their development, even to the point that the genital tubercle becomes a fully formed penis instead of a clitoris.

#### Principle of the critical period

Both of the foregoing examples of an error in embryonic sexual development illustrate the principle of the critical period, that is to say, that period in the development of some system or function when the organism passes from a neutral or undifferentiated stage to a differentiated one. It is during this critical period that development is vulnerable to interference and deflection, which, if they occur, will leave a permanent residual in structure or function. The further development of the organism will then in some degree be determined by, or influenced by, the outcome of the critical period.

Effects of atypical early experiences. The concept of the critical period applies not only to morphologic and prenatal development, but to behavioral, psychological, and postnatal development as well. In behavioral science, the European animal ethologists were the first to observe that patterns of behavior, including sexual behavior, can be experimentally changed during a critical, developmental learning period, and thenceforth



this deviance becomes relatively fixed. In other words, particular atypical experiences may interfere with or deform the genetic, hormonal, or neural norms of behavioral development otherwise expected.

The justly famous experiments of Harlow (see Money 1965a, chapter 7) on the macaque monkey are a case in point. The experimental monkeys showed that opportunity to play with age-mates in early childhood was a developmental prerequisite to the achievement of normal copulatory ability in adolescence. Depriving the infants of childhood play, combined with depriving them of interaction with their mothers by rearing them with dummy mothers made of wire covered with a piece of towel, had particularly adverse effects. Such doubly deprived animals were unable to adapt to the motions of mating and to effect sexual intercourse even when paired with mates selected for experience and gentleness. Those few females in whom pregnancy was finally achieved were incompetent to mother their own young. They neglected and injured them, sometimes cruelly. Other aspects of their behavior, in general, also appeared to be grossly disordered.

Another illustrative experiment is that of Birch (1956). Female rats were raised in isolation with a rubber collar or ruff around their necks, which deprived them of the experience of licking themselves or others. They were finally released from their collars to deliver their first litter. Instead of licking the young as they were born, they ate most of them. The surviving 5 per cent were retardedly and improperly retrieved and badly suckled; some were eaten after being carried to the nest. Three survivors of a large litter died of starvation. Their mother had shepherded them under her chin every time they approached her belly to search for a nipple.

Experiments in population density with rats have demonstrated yet another way in which behavioral experience, namely crowding, may induce aberrant sexual and social behavior (Calhoun 1962), affecting even the unborn fetuses and their future behavior, presumably via the maternal endocrine system (Keeley 1962).

The proper understanding of psychosexual development and of erroneous development in human beings requires focusing attention on experience, behavior, and social interaction during the formative years of childhood. In man, perhaps even more so than in lower species, peculiarities and special features of experience during critical periods may change the expected genetic, hormonal, or neural norms of development.

**Hermaphroditism—an example.** Probably the most dramatic illustration of this last proposition is to be found in the psychology of hermaphroditism (Money, Hampson, & Hampson 1957; Money 1961). Human hermaphroditism, for which there are several different etiologies, is conventionally defined in terms of a discrepancy between the gonads (ovaries, testes, or, rarely, ovotestes) and the morphology of the external genitalia. Sometimes the discrepancy may be total, as indicated at the beginning of this article, namely, when the gonads are testicular and the external organs are perfectly female (the syndrome of testicular feminization) or when the external organs are perfectly male except for an empty scrotum, the gonads being ovaries and in proper ovarian position (a rare variant of the hyperadrenocortical syndrome of female hermaphroditism). Most commonly in hermaphroditism, however, the discrepancy between gonads and external organs is incomplete, by reason of the fact that the external organs are themselves imperfectly differentiated as either male or female. In this incompletely state of differentiation, the external sex organs of either sex look remarkably similar: the penis with a urinary gutter instead of a tube and with its orifice near the scrotum can easily be confused with an enlarged clitoris, and the incompletely fused scrotum may pass for labia that have improperly begun to fuse. Because confusion is possible and because medical decisions may sometimes differ, it may so happen that two individuals of identical genetic, gonadal, and hormonal diagnosis are assigned to different sexes.

**Effects of sex assignment.** The expected outcome of such sex assignment is that psychosexual differentiation will proceed to take place congruously with assignment and rearing. Ideally, the external organs will have been repaired surgically as early as possible, so as to conform to the assigned sex. The visible, kinesthetic, and tactile body image is a potent fact in a child's own awareness of his or her gender identity, and the visible appearance of sex is, of course, the ultimate criterion for the gender-role expectancies imposed by most of the people who are significant in a child's life.

Even with delayed genital repair, it has proved possible for psychosexual differentiation to take place in keeping with assigned sex when genetic, gonadal, and hormonal sex are at variance. Incongruous-looking sex organs militate against psychosexual differentiation appropriate to sex assignment, especially when the incongruity is reinforced by some other source of ambiguity such as paren-

tal lack of conviction as to the way they should be rearing their child or open doubt expressed by age-mates. In some instances, doubt may develop in the child's own mind without any clear retrospective indication of its origin.

Ideally, for psychosexual differentiation to be brought to maturity at puberty, assigned sex should not be at variance with hormonal sex and the secondary sexual changes of the body that are hormonally controlled. This ideal can be met in modern endocrinology. Yet, there are occasional untreated cases; among them are cases of hormonal virilization in girls with the hyperadrenocortical syndrome, which is remarkable in view of the intensity of virilization and its early onset before the usual age of puberty. These cases show that the persistence of a feminine psychosexual differentiation is possible despite the hormonal contradiction and its severe handicap to proper adolescent social maturation.

**Crucial variables.** In psychosexual differentiation, the crucial variables, which may in hermaphroditism be independent of one another and may be overridden by the experiential effects of assignment and rearing, are five: (1) chromosomal sex, as determined by actual chromosomal count, (2) gonadal sex, (3) hormonal sex, (4) internal accessory morphologic sex, and (5) external morphologic sex. Psychosexual differentiation itself transcends not only all five of these variables but also the assigned sex in the syndrome of transvestism with transsexualism, which has, therefore, been treated, and not without reasonable success, by surgical, hormonal, and social sex reassignment (Benjamin 1964; Pauly 1965).

To have found rare clinical cases in which the five physical variables can be overridden is not, of course, to have proved that the physical variables do not under other circumstances contribute to, or help determine, psychosexual differentiation. What has been proved is not that psychosexual differentiation is always and exclusively a matter of behavioral and experiential life history but that experiences determined by rearing are far more potent than might otherwise have been expected. Harlow's investigations on the macaque demonstrated, on the one hand, the extraordinary power and permanence of the social-experience factor in preventing psychosexual maturity; but they also led him, on the other hand, to conclude that sex differences in normal macaque behavior antedate childhood experience and learning.

Harlow observed (1962) that the young male monkeys made more threats toward other monkeys, both male and female, than did the young females.

The females' threats, moreover, were reserved primarily for other young females. The young females retreated more often than the males, specifically by adopting the female sexual posture. The young males initiated more play contacts, with playmates of either sex, than did the females; and the males had a monopoly on rough-and-tumble play. With increasing age, the males showed increasing frequency of the male mounting position in their copulatory play. They showed little grooming behavior, which, in adults, is more a characteristic of female than male sexual behavior.

The application of Harlow's method of direct observations and systematic experiments on monkeys is restricted in human beings, as compared with clinical case study, by reason of the usual ethical vetoes on experimental manipulation of the lives of human beings and by the customary taboos imposed on sexual play and the expression of sexual interests in childhood. Systematic experimental attempts to study the development of masculinity and femininity have confined themselves to the secondary data of questionnaires or interviews or to the oblique data of free association, projective tests, and play. Although not thoroughly consistent or definitive, the findings show that psychosexual differentiation varies with age and in some degree is related to family and class differences in child-rearing practices and the emotional quality of the relationship between the child and each parent.

**Sex reassignment.** That there is a critical period in psychosexual differentiation, and that it is early in life, is suggested by clinical evidence, again from hermaphroditism, namely in instances of a change of the sex of assignment (Money, Hampson, & Hampson 1956). It is not difficult for a child under a year of age to adjust to a sex reassignment, provided the parents can also adjust. But the older the child—for example, if he is past the age of learning to talk and is becoming acquainted with the verbal and other signals of gender difference—and the further that psychosexual differentiation has progressed without ambiguity, the progressively less likely it is that an imposed sex reassignment will be successfully assimilated. It thus appears that the establishment of a psychosexual identity and gender role roughly parallels the learning of the native language. The critical period, in which the basic essentials are laid down, is ended probably by the kindergarten age. The remaining years of childhood are years of consolidation until the hormonal changes of adolescence bring psychosexual differentiation to its completed expression in both the subjective









tropins cyclically, as in the female; and it allows mating pattern elements that are chiefly masculine, although they are shared by both sexes, to predominate in childhood.

### Some dimensions of deviance

**Androgen as an organizer.** Androgen is thus implicated as an active organizer of the sexual control system, both in terms of reproductive cycles and sexual behavior. At a critical period earlier in embryonic life, androgen is an active organizer substance without which the internal anlagen of the female reproductive anatomy persist and the anlagen of the external organs feminize (Jost 1958). Nature's primary impulse, in other words, is to make a female. One conjectures a relationship here with the greater vulnerability of the male to disease and death, which would create an unequal sex ratio except for the fact that 106 males are born to 100 females. One conjectures also a connection with the fact that psychosexual disorders have a higher incidence in men than in women, with some of the psychosexual anomalies, like fetishism, voyeurism, and exhibitionism, being to all intents and purposes unheard of in females.

**Sex differences in arousal patterns.** The higher incidence of psychosexual pathologies in the male may be related to a sex difference in erotic arousal patterns. This difference is manifested in the greater capacity of the male to be aroused erotically and genitopelvicly, and to be ready for immediate release, by pictorial and narrative imagery. Such stimuli arouse the female sentimentally to want her lover or husband; but she is dependent on direct tactile stimulation, more than the male, for the arousal of genitopelvic eroticism.

There is some evidence in the psychology of hermaphroditism that the male arousal pattern may be neurally organized under the influence of androgen in fetal life. It will be for further research to determine whether the development of psychosexual pathology in the male may be rooted in part in imperfect or faulty neurohormonal organization at this early period. One thing already quite clear in the evidence of hermaphroditism is that a masculine type of arousal pattern is perfectly compatible with entirely feminine cognitive content and imagery of eroticism that parallels the sex of assignment.

**Arousal stimuli.** Perceptual stimulants to sexual arousal and the reproductive cycle vary with the species. The relation between light stimulation and the release of pineal hormones is being newly investigated in the rat. Olfactory stimuli also are

the subject of important new experiments (Parkes & Bruce 1961), which show, after copulation in the mouse, that odors released by the male regulate the establishment of pregnancy in the female. Acuity of the sense of smell in human females (Money 1965b) is regulated by estrogen and varies with the menstrual cycle. Its relation to other cyclic sexual phenomena or to sexual disorder has not been investigated. Odor may be the agent responsible, in the macaque monkey, for fluctuation in male sexual behavior in rhythm with the female menstrual cycle, as described by Michael and Herbert (1963).

In addition to benefiting from the new neurophysiological techniques, contemporary sex research has also benefited from the new techniques in genetics. These techniques have enabled researchers to establish the presence (female) or absence (male) of the sex-chromatin mass (Barr body) in the nucleus of the body's cells and to photograph, measure, and count the actual chromosomes in cells. The relevance of the new findings to the theory of psychosexual pathology has been chiefly negative: gross chromosome defects are not demonstrable in the majority of psychosexual disorders. A partial exception may exist in the case of Klinefelter's syndrome. Affected individuals are phenotypic males with an extra X chromosome. There appears to be an unusually high incidence of psychopathology, including sexual disorders, that develops in association with the syndrome (Money 1963a). Among the sexual disorders are homosexuality and transvestism (Money & Pollitt 1964). Otherwise genetics cannot be directly implicated in psychosexual ambiguity (Money 1963b).

Disorders of sexual psychology express themselves in full in adolescence after the advent of hormonal puberty brings erotic functioning to maturity. It is probable that the majority of such disorders have their initial origins in faulty psychosexual differentiation persisting from early childhood. Later error or regression of differentiation are also possible, as is simple arrestment in psychosexual growth.

Faults, errors, regression, and arrest of development have, on the one hand, a chronologic origin and, on the other hand, an etiologic origin. The chronologic origin may fall within a critical period, with resultant immutability or refractoriness to change. The etiologic origin may be single or multiple. An etiologic factor may by itself represent only a potential vulnerability to psychosexual disorder, materializing only when in combination

with another factor or factors. Thus a genetic vulnerability may materialize only when the behavioral environment conspires appropriately. Etiologic considerations include genetic, gestational, metabolic, nutritional, toxic-infective, traumatic, behavioral-experiential, and social-interactional factors.

JOHN MONEY

[Other relevant material may be found in AFFECTION; DEVELOPMENTAL PSYCHOLOGY; GENETICS; IDENTITY; PSYCHOSOCIAL; INFANCY, article on THE EFFECTS OF EARLY EXPERIENCE; NERVOUS SYSTEM; PSYCHOANALYSIS; SOCIALIZATION.]

#### BIBLIOGRAPHY

- BENJAMIN, H. 1964 Nature and Management of Transsexualism: With a Report on Thirty-one Operated Cases. *Western Journal of Surgery, Obstetrics and Gynecology* 72:105-111.
- BIRCH, H. G. 1956 Sources of Order in the Maternal Behavior of Animals. *American Journal of Orthopsychiatry* 26:279-284.
- CALHOUN, JOHN B. 1962 Population Density and Social Pathology. *Scientific American* 206, Feb.:139-148.
- FISHER, ALAN E. 1956 Maternal and Sexual Behavior Induced by Intracranial Chemical Stimulation. *Science* 124:228-229.
- GLASCOCK, R. F.; and MICHAEL, R. P. 1962 The Localization of Oestrogen in a Neurological System in the Brain of the Female Cat. *Journal of Physiology* 163: 38P-39P.
- GREEN, RICHARD; and MONEY, JOHN 1961 Effeminacy in Prepubertal Boys: Summary of Eleven Cases and Recommendations for Case Management. *Pediatrics* 27:286-291.
- HARLOW, HARRY F. 1962 The Heterosexual Affectional System in Monkeys. *American Psychologist* 17:1-9.
- HARRIS, G. W.; MICHAEL, R. P.; and SCOTT, P. P. 1958 Neurological Site of Action of Stilboestrol in Eliciting Sexual Behavior. Pages 236-254 in Ciba Foundation, *Symposium on the Neurological Basis of Behavior*. Edited by G. E. W. Wolstenholme and C. M. O'Connor. Boston: Little. → Includes two pages of discussion.
- JOST, ALFRED 1958 Embryonic Sexual Differentiation: Morphology, Physiology, Abnormalities. Pages 15-45 in Howard W. Jones and William W. Scott (editors), *Hermaphroditism, Genital Anomalies and Related Endocrine Disorders*. Baltimore: Williams & Wilkins.
- KAWAKAMI, M.; and SAWYER, CHARLES H. 1959 Neuroendocrine Correlates of Changes in Brain Activity Thresholds by Sex Steroids and Pituitary Hormones. *Endocrinology* 65:652-668.
- KEELEY, KIM 1962 Prenatal Influence on Behavior of Offspring of Crowded Mice. *Science* 135:44-45.
- MACLEAN, PAUL D. 1962 New Findings Relevant to the Evolution of Psychosexual Functions of the Brain. *Journal of Nervous and Mental Disease* 135:289-301.
- MACLEAN, PAUL D.; DUA, S.; and DENNISTON, R. H. 1962 A Mapping of Cerebral Structures Involved in Seminal Discharge and Genital Scratching. American Neurological Association, *Transactions* 87:136-139.
- MICHAEL, RICHARD P. 1961 An Investigation of the Sensitivity of Circumscribed Neurological Areas to Hormonal Stimulation by Means of the Application of Oestrogens Directly to the Brain of the Cat. Pages 465-480 in International Neurochemical Symposium, Fourth, Varenna, Italy, 1960, *Regional Neurochemistry . . . Proceedings*. Oxford: Pergamon Press.
- MICHAEL, RICHARD P. 1962 Estrogen-sensitive Neurons and Sexual Behavior in Female Cats. *Science* 136: 322-323.
- MICHAEL, RICHARD P.; and HERBERT, J. 1963 Menstrual Cycle Influences Grooming Behavior and Sexual Activity in the Rhesus Monkey. *Science* 140:500-501.
- MONEY, JOHN 1961 Hermaphroditism. Volume 1, pages 472-484 in Albert Ellis and Albert Abarbanel (editors), *Encyclopedia of Sexual Behavior*. New York: Hawthorn.
- MONEY, JOHN 1963a Cytogenetic and Psychosexual Incongruities With a Note on Space-Form Blindness. *American Journal of Psychiatry* 119:820-827.
- MONEY, JOHN 1963b Factors in the Genesis of Homosexuality. Pages 19-43 in Conference on Community Mental Health Research, Fourth, Washington University, St. Louis, 1962, *Determinants of Human Sexual Behavior: Proceedings*. Springfield, Ill.: Thomas.
- MONEY, JOHN 1963c Psychosexual Development in Man. Volume 5, pages 1678-1709 in *Encyclopedia of Mental Health*. New York: Watts.
- MONEY, JOHN (editor) 1965a *Sex Research: New Developments*. New York: Holt. → See especially Chapter 7, "The Effect of Rearing Conditions on Behavior," by Harry F. Harlow and M. K. Harlow.
- MONEY, JOHN 1965b Influence of Hormones on Sexual Behavior. *Annual Review of Medicine* 16:67-82.
- MONEY, JOHN; HAMPSON, J. G.; and HAMPSON, J. L. 1956 Sexual Incongruities and Psychopathology: The Evidence of Human Hermaphroditism. *Johns Hopkins Hospital Bulletin* 98:43-57.
- MONEY, JOHN; HAMPSON, J. G.; and HAMPSON, J. L. 1957 Imprinting and the Establishment of Gender Role. *A.M.A. Archives of Neurology and Psychiatry* 77:333-336.
- MONEY, JOHN; and POLLITT, ERNESTO 1964 Cytogenetic and Psychosexual Ambiguity: Klinefelter's Syndrome and Transvestism Compared. *Archives of General Psychiatry* 11:589-595.
- OLDS, JAMES 1958 Self-stimulation Experiments and Differentiated Reward Systems. Pages 671-687 in Herbert H. Jasper et al. (editors), *Reticular Formation of the Brain*. Boston: Little.
- PARKES, A. S.; and BRUCE, H. M. 1961 Olfactory Stimuli in Mammalian Reproduction. *Science* 134:1049-1054.
- PAULY, IRA B. 1965 Male Psychosexual Inversion: Transsexualism. *Archives of General Psychiatry* 13:172-181.
- PRITCHARD, MICHAEL 1962 Homosexuality and Genetic Sex. *Journal of Mental Science* 108:616-623.
- SOCIETY OF MEDICAL PSYCHOANALYSTS 1962 *Homosexuality: A Psychoanalytic Study*, by Irving Bieber et al. New York: Basic Books.

#### IV

#### SEXUAL DEVIATION: SOCIAL ASPECTS

Deviance in general, and sexual deviance in particular, presents the scientist with extremely difficult problems of definition and analysis. Sexuality is, perhaps more than any other aspect of human behavior, mingled with moral imperatives, conscious fantasy, and unconscious desires, all of which combine to trouble even most scientific ob-









against which strong sanctions are invoked, that does not seem to be learned in recognizable deviant social structures and that does not have a socially organized performance system.

**Functional deviance.** Masturbation and premarital coitus are primary examples of behavior that is generally defined as deviant but that does not involve the application of sanctions unless the perpetrators are, by some mischance, apprehended or there is some untoward consequence, such as pregnancy or venereal disease. These forms of behavior may be described as "wrong but normal." There is considerable class variation in the incidence and frequency of masturbation and premarital coitus, with the former important among middle-class males and the latter among lower-class males. Solitary masturbation is not against the law; premarital coitus may be, especially if either party is legally a minor, but even so prosecutions are comparatively rare.

The perceived utility of masturbation as a male adolescent sexual outlet with few social consequences in comparison to coitus may account for the decreasing stigma attached to it. It seems to be related to other behavior defined by social class and characterizes males who assume middle-class occupational roles (Kinsey et al. 1948, chapter 10). Premarital coitus is culturally valued for males especially within the pro-masculine lower-class culture, accounting for its prevalence and the lack of invoked sanctions (I. Reiss 1960; Whyte 1943). Both masturbation and coitus are class-linked and seem to be the most appropriate responses to the impact of the sexual drive after puberty in the male. For females premarital coitus is connected with impending marriage and a love relationship; thus, the behavior is defined as proper in a situation-specific way. Masturbation in females seems to have no class linkages at all. In general these activities have a smaller incidence and a lower frequency among females than among males, possibly because of both hormonal differences and differences in socialization.

**Deviant subcultures.** The second type of deviant sexual behavior is that which involves supportive social structures either for the purpose of entry into the system of behavior or for the maintenance of the individual in the behavior. This type of behavior comprises homosexuality and prostitution. It is only female prostitution that has all the elements of a deviant subculture similar to that of drug addicts or juvenile delinquents (Davis 1961; Kinsey et al. 1948, chapter 10). In this situation the stigma for the sexual behavior is attached to the prostitute, and the legal sanctions for

the behavior are leveled primarily against her, secondarily against the abetting persons (pimps, madams), and exceedingly rarely against the customers or against agents of law enforcement who may condone her activities. Prostitution develops organized forms primarily because of its commercial aspect and because of the character of the enforcement program that is designed to control it.

For the most part, recruitment of individuals for homosexual experience does not occur through a social system organized for that purpose. The majority of persons who come to conceive of themselves as homosexual do so on the basis of readying experiences early in life that are linked to secondary reinforcements in puberty. Unfortunately, studies of the etiology of homosexuality have been most unsatisfactory in their attempts to locate the crucial early experiences that are salient (Money 1963). Some etiological studies have focused on biological and especially genetic factors, sometimes mediated through hormonal anomalies. But even though there have been some studies of twins (Kallman 1952; 1963) and some work has been done on hormonal factors (Kinsey 1941), the most persuasive discussions of early factors in the development of a homosexual commitment have been couched in the psychoanalytic tradition or in certain variants of it (Society . . . 1962; Ovesey 1954). The psychoanalytic tradition locates the origins of homosexuality in pathological versions of parent-child relations in which the mother provokes her son's sexual interest and in consequence burdens him with an anxiety about incest with which he cannot deal. As a result his relations with females are blighted because of his fears of paternal revenge for his incest, which remains forever unconsummated. Other explanations in this tradition are less metaphorical in their language but still focus upon seductive mothers, detached fathers, lack of a male figure in the home, and the like. Unfortunately the bulk of the populations who have been the source of the data in these reports have been examined in the course of psychotherapy or other treatment, creating a situation in which pathology and homosexuality are inextricably confounded.

Psychological research into homosexuality has also maintained a determined concern with the etiological question and with developing instrumentation that might successfully distinguish between the homosexual and the heterosexual on the basis of either manifest items or latent items that are presumed to be indicators of homosexual commitments (see, for instance, the psychometric indices of masculinity and femininity used by Brown 1958;

Hooker 1957; 1958; Dean & Richardson 1964; Freund 1963).

During the pubescent years there is a good deal of minor homosexual experimentation that accounts for nearly all of the homosexual activity of the predominantly heterosexual portion of the population. Thus, it can be predicted that about one in three of all males born in the United States will at some time have a homosexual experience to the point of orgasm; however, about 85 per cent of those having such experience will do so only in their adolescent years. Among females the proportion who ever have a homosexual experience will be somewhat smaller, but once again most of them will confine the experience to adolescence (Kinsey et al. 1948; 1953). The factors that distinguish those who go on to adult homosexual commitments from those who experience homosexual acts only as transient parts of adolescent development are still unknown. However, it is known that males often develop strong sexual interests in other males without knowing that there is such a thing as homosexuality. (This process occurs more frequently in specialized sexual interests such as transvestism and fetishism.)

The realization of being homosexual occurs in a variety of ways: sometimes through processes of introspection, sometimes through falling in with other homosexuals who act seductively, and sometimes through sexual contacts with other males who finally become defined as homosexual. It is after the point of conscious admission that one is a homosexual—an admission that ranges from the traumatic to the delighted—that the search for sexual partners begins. This search usually requires entry into the world of homosexual bars, development of the argot of homosexual life, and gradual learning of the modes of approach and retreat that are related to the satisfaction of sexual needs. Study of this zone of the homosexual's adjustment is beginning to result in a literature that suggests that adult homosexual commitments are not re-enactments of early pathological relations between parents and children. These studies have begun to outline the processes of adult socialization that are involved in the development of the homosexual commitment. There now also exists a literature on particular aspects of the homosexual life (Hooker 1966; Leznoff & Westly 1956; Achilles 1967).

The extant literature is nearly exclusively concerned with the male homosexual; the literature on the lesbian is both scanty and scientifically inadequate. In general, discussions of the lesbian tend to focus on the similarities between male and

female homosexuality and to see one as the mirror image of the other—that is, to see male homosexuals as imitations of females and female homosexuals as imitations of males. A more satisfactory view, however, is to see female homosexuality as continuous with normal female sexuality (Simon & Gagnon 1967; Ovesey 1956).

Not all homosexuals are regular members of the homosexual community and many of them never involve themselves in this social milieu at all. It is difficult to determine whether or not a more intensive or a more detached relationship to the community is more helpful in the homosexual's personal and social adjustment. The maintenance of female homosexuality occurs more often outside the community context than does homosexuality among males (Simon & Gagnon 1967).

The development of homosexual contacts among prisoners in penal institutions is more like an articulated system of introduction to homosexual experience than is homosexual development in the free community. Homosexual behavior in prisons ranges from the careful seduction, in adult male institutions, of a weaker male by an aggressive male (involving money and protection from the latter in exchange for the sexual cooperation of the former, and often leading to "marriage") to the "extended family" systems observed in juvenile female institutions and the homosexual "bringing out" in female penitentiaries (Lindner 1948; Sykes 1958; Halleck & Hersko 1962; Lamberti 1963; Ward & Kassebaum 1965; Giallombardo 1966).

**The antisocial deviant.** The third type of deviant sexual acts to be considered are those most specifically prosecuted by law-enforcement agencies; they involve aggression, youth, close kin relations, and public disturbance (a typology suggested both in Wheeler 1960 and in Indiana University 1965; see pp. 1-13 for the classification of sex offenders). Most of these acts are committed by individuals whose behavior is the result of needs developed outside any subcultural or deviant training ground composed of other persons oriented toward the same behavior. The origins of heterosexual pedophilia, for example, seem to lie in the incapacity of the male to establish any meaningful relationship with an adult female or in many cases a failure to find any sociosexual outlet at any level of psychological involvement. The pedophile resorts to children because the difference in age gives him a sense of power over them as well as the hope that he will meet with an uncritical reception. There is no social support for the act either before or after its occurrence. The individual has the latent capacity to commit the act without any



social support, and at the intersection of the appropriate forces (aging, alcohol, fatigue, emotional stress) a sexual contact with a child occurs. The model of the behavior that seems to be most useful is that of the releasing mechanism. The pedophile has learned in childhood a certain set of responses that gives him the potential for the act as an adult. Under the appropriate circumstances the potential is realized.

Other sexual acts of the third type that have some of the aspects of pedophilia are homosexual pedophilia, exhibitionism, compulsive peeping, and sex offenses involving incest and the use of force. In the cases of peeping, assaultive offenses, and incest there are certain complicating social or cultural factors. Peeping and the use of minimal aggression in sexual relations are not necessarily viewed as abnormal or deviant. The act of peeping is something that many males have performed when confronted by a lighted unshaded window; however, for the impulse to invade and dominate the sexual life of the individual to such an extent that he must seek out windows in which to look suggests that more than a simple heterosexual interest is involved. A certain amount of aggression by males is expected in the process of sexual exploration in Western culture, and it is clear that excessive force is used more frequently than is reported to the agencies of law enforcement. This is predominantly a lower-class phenomenon, but not totally so, since there is evidence for its occurring among college-level males as well (Kirkpatrick & Kanin 1957). There are indications that those college-level males who do use force have certain other characteristics that differentiate them from males who do not use force. The aggressive male is commonly responding to his own needs rather than to the interactional situation in which he is involved. What may be involved is a blunting of the capacity to recognize negative signs displayed by the female, either as a function of the male's specific personality characteristics or as a function of his sexual excitement.

*Incest.* The problem of sexual contact between relatives is confined primarily to father-daughter contacts and to a lesser extent to those between brothers and sisters and between cousins. The last two are usually transitory childhood contacts or those that take place just after puberty, and in the case of brother-sister incest they may be made more complicated by the presence of mental deficiency. Father-daughter contacts are most prevalent in rural, or rural-origin, families. The families in which such contacts occur commonly are extremely disorganized families in which the father

converts the daughter into a source of sexual and often pseudo-marital relations (Kaufman et al. 1954). Mothers in these families retreat from the center of the family life, and only if the daughter is quite young does the mother inform the police. With very young daughters the offense often has elements of pedophilia.

*Sex-offense rates.* A useful point to be made about deviant behavior of this third type is that over long periods of time it represents a relatively stable proportion of all offenses known to the police, of all persons arrested, and of all felons in penal institutions. In addition, the rates per hundred thousand in the population have remained stable for many years. The stability of these rates indicates that disturbed and disorganized families represent a rather constant source of individuals who grow up with strategic weaknesses in their character structures and thus become sexual offenders. The rate of increase in the number of such persons, however, does not exceed the growth rate of the population, indicating that the roots of these kinds of deviant acts are in the family, as the basic unit of social organization, rather than in secondary social structures.

As can be seen from the foregoing discussion, the forms of sexual behavior that are approved arise from sources very similar to those from which the forms called deviant arise, and the margins between one and the other are sharp in moral evaluation but vague in fact. The general situation of sexuality in this society may be best described as that of pluralistic ignorance, a condition of noninformation and misinformation that serves to direct (or misdirect) a powerful drive.

For the vast majority of deviant sexual acts, negative sanctions are frequently expressed but only sporadically enforced. For a smaller proportion of acts, deviant behavior involves recognizable social organization that serves either as a learning structure or as a performance structure for the development and maintenance of the behavior. Finally, for a very small proportion there seem to be no social structural components to the behavior, either for its development or for its continuation, and its origins seem most likely to be in early familial life, as was explained above.

With the general absence of mediating social structures between the original development of sexual capacity in childhood and the performance of sexual roles in adulthood the conditions described above are not surprising. Sexual roles are worked out on an individual basis with a general incapacity to share and reformulate very large portions of the

sexual experience. Since there is a high premium on silence in sexual matters, there is rarely an opportunity for social checks on deviant impulses, nor does there exist any system by which the behavior of individuals, if it varies from the rigid normative standards, can be examined in terms of the behavior's relevant social and psychological meaning. Of all of the areas of deviant behavior, it is sexual deviance which most strongly maintains its "individualistic" character.

JOHN H. GAGNON

[See also DEVIANT BEHAVIOR; MENTAL DISORDERS; and the biographies of FREUD; KINSEY.]

#### BIBLIOGRAPHY

- ACHILLES, NANCY 1967 The Development of the Homosexual Bar as an Institution. Pages 214-228 in John H. Gagnon and William Simon (editors), *Sexual Deviance*. New York: Harper.
- BELL, ROBERT R.; and BUERKLE, JACK V. 1961 Mother and Daughter Attitudes to Premarital Sexual Behavior. *Marriage and Family Living* 23:390-392.
- BROWN, DANIEL G. 1958 Inversion and Homosexuality. *American Journal of Orthopsychiatry* 28:424-429.
- DAVIS, KINGSLEY 1961 Prostitution. Pages 262-288 in Robert K. Merton and Robert A. Nisbet (editors), *Contemporary Social Problems*. New York: Harcourt.
- DEAN, ROBERT B.; and RICHARDSON, HAROLD 1964 Analysis of MMPI Profiles of Forty College-educated Overt Male Homosexuals. *Journal of Consulting Psychology* 28:483-486.
- EHRMANN, WINSTON 1959 *Premarital Dating Behavior*. New York: Holt.
- FORD, CLELLAN S.; and BEACH, FRANK A. 1951 *Patterns of Sexual Behavior*. New York: Harper.
- FREUD, SIGMUND (1905) 1953 Three Essays on the Theory of Sexuality. Volume 7, pages 123-245 in *The Standard Edition of the Complete Psychological Works of Sigmund Freud*. London: Hogarth; New York: Macmillan. → First published as *Drei Abhandlungen zur Sexualtheorie*.
- FREUND, KURT 1963 A Laboratory Method for Diagnosing Predominance of Homo- or Hetero-erotic Interest in the Male. *Behavior Research and Therapy* 1:85-93.
- GAGNON, JOHN H.; and SIMON, WILLIAM (editors) 1967 *Sexual Deviance*. New York: Harper.
- GIALLOMBARDO, ROSE 1966 *Society of Women: A Study of a Woman's Prison*. New York: Wiley.
- HALLECK, SEYMOUR L.; and HERSKO, MARVIN 1962 Homosexual Behavior in a Correctional Institution for Adolescent Girls. *American Journal of Orthopsychiatry* 32:911-917.
- HAMPSON, JOHN L.; and HAMPSON, JOAN G. 1961 The Ontogenesis of Sexual Behavior in Man. Pages 1401-1432 in William C. Young (editor), *Sex and Internal Secretions*. 3d ed. 2 vols. Baltimore: Williams & Wilkins.
- HARLOW, HARRY F. 1959 Love in Infant Monkeys. *Scientific American* 200, no. 6:68-74.
- HOOKE, EVELYN 1957 The Adjustment of the Male Overt Homosexual. *Journal of Projective Techniques* 21:18-31.
- HOOKE, EVELYN 1958 Male Homosexuality in the Rorschach. *Journal of Projective Techniques* 22:33-54.
- HOOKE, EVELYN 1966 The Homosexual Community. Pages 354-364 in James O. Palmer and Michael J. Goldstein (editors), *Perspectives in Psychopathology*. New York: Oxford Univ. Press.
- INDIANA UNIVERSITY, INSTITUTE FOR SEX RESEARCH 1965 *Sex Offenders: An Analysis of Types*, by Paul H. Gebhard, J. H. Gagnon, W. B. Pomeroy, and C. V. Christenson. New York: Harper.
- KALLMAN, FRANZ J. 1952 Comparative Twin Study on the Genetic Aspects of Male Homosexuality. *Journal of Nervous and Mental Disease* 115:283-298.
- KALLMAN, F. J. 1963 Genetic Aspects of Sex Determination and Sexual Maturation Potentials in Man. Pages 5-18 in Conference on Community Mental Health Research, 4th, Washington University, St. Louis, 1962, *Determinants of Human Sexual Behavior*. Edited by George Winokur. Springfield, Ill.: Thomas.
- KARPMAN, BENJAMIN 1954 *The Sexual Offender and His Offenses*. New York: Julian Press.
- KAUFMAN, IRVING; PECK, A. L.; and TAGIURI, C. K. 1954 The Family Constellation and Overt Incestuous Relations Between Father and Daughter. *American Journal of Orthopsychiatry* 24:266-277.
- KINSEY, ALFRED C. 1941 Criteria for a Hormonal Explanation of the Homosexual. *Journal of Clinical Endocrinology* 1:424-428.
- KINSEY, ALFRED C.; POMEROY, W. B.; and MARTIN, C. E. 1948 *Sexual Behavior in the Human Male*. Philadelphia: Saunders.
- KINSEY, ALFRED C. et al. 1953 *Sexual Behavior in the Human Female*. Philadelphia: Saunders.
- KIRKPATRICK, CLIFFORD; and KANIN, EUGENE 1957 Male Sex Aggression on a University Campus. *American Sociological Review* 22:52-58.
- LAMBERTI, ARMAND 1963 A Type of Peer Relationship in a Girls' Training School. *Menninger Clinic, Bulletin* 27:200-204.
- LEZNOFF, MAURICE; and WESTLEY, WILLIAM 1956 The Homosexual Community. *Social Problems* 3:257-263.
- LINDNER, ROBERT M. 1948 Sexual Behavior in Penal Institutions. Pages 201-215 in Albert Deutsch (editor), *Sex Habits of American Men: A Symposium on the Kinsey Report*. Englewood Cliffs, N.J.: Prentice-Hall.
- MARMOR, JUDD (editor) 1965 *Sexual Inversion: The Multiple Roots of Homosexuality*. New York: Basic Books.
- MONEY, JOHN 1961 Sex Hormones and Other Variables in Human Eroticism. Pages 1383-1400 in William C. Young (editor), *Sex and Internal Secretions*. 2 vols. 3d ed. Baltimore: Williams & Wilkins.
- MONEY, JOHN 1963 Factors in the Genesis of Homosexuality. Pages 19-43 in Conference on Community Mental Health Research, 4th, Washington University, St. Louis, 1962, *Determinants of Human Sexual Behavior*. Edited by George Winokur. Springfield, Ill.: Thomas.
- OVESEY, LIONEL 1954 The Homosexual Conflict: An Adaptational Analysis. *Psychiatry* 17:243-250.
- OVESEY, LIONEL 1956 Masculine Aspirations in Women: An Adaptational Analysis. *Psychiatry* 19:341-351.
- REISS, ALBERT J. JR. 1960 Sex Offenses: The Marginal Status of the Adolescent. *Law and Contemporary Problems* 25:309-333.
- REISS, ALBERT J. JR. 1961 The Social Integration of Queers and Peers. *Social Problems* 9:102-120.
- REISS, IRA L. 1960 *Premarital Sexual Standards in America*. Glencoe, Ill.: Free Press.



- RIEFF, PHILIP 1959 *Freud: The Mind of the Moralist*. New York: Viking.
- SEARS, ROBERT R.; MACCOBY, E. E.; and LEVIN, H. 1957 *Patterns of Child Rearing*. Evanston, Ill.: Row, Peterson.
- SIMON, WILLIAM; and GAGNON, JOHN H. 1967 The Lesbians: A Preliminary Overview. Pages 247-282 in John H. Gagnon and William Simon (editors), *Sexual Deviance*. New York: Harper.
- SOCIETY OF MEDICAL PSYCHOANALYSTS 1962 *Homosexuality: A Psychoanalytic Study*, by Irving Bieber et al. New York: Basic Books.
- SPITZ, RENÉ A. 1949 Autoeroticism. *Psychoanalytic Study of the Child* 3/4:85-120.
- SUTHERLAND, EDWIN H.; and CRESSEY, DONALD R. (1924) 1960 *Principles of Criminology*. 6th ed. Philadelphia: Lippincott. → First published as a textbook under the title *Criminology*, with Sutherland as the sole author.
- SYKES, GRESHAM M. 1958 *The Society of Captives: A Study of a Maximum Security Prison*. Princeton Univ. Press.
- TRILLING, LIONEL (1950) 1957 The Kinsey Report. Pages 223-242 in Lionel Trilling, *The Liberal Imagination*. Garden City, N.Y.: Doubleday.
- WARD, DAVID A.; and KASSEBAUM, GENE B. 1965 *Women's Prison: Sex and Social Structure*. Chicago: Aldine.
- WHEELER, STANTON 1960 Sex Offenses: A Sociological Critique. *Law and Contemporary Problems* 25:258-278.
- WHYTE, WILLIAM F. 1943 A Slum Sex Code. *American Journal of Sociology* 49:24-31.
- YOUNG, WILLIAM C. (editor) 1961 *Sex and Internal Secretions*. 2 vols. 3d ed. Baltimore: Williams & Wilkins. → Pagination is continuous.
- YOUNG, WILLIAM C.; GOY, ROBERT W.; and PHOENIX, CHARLES H. 1964 Hormones and Sexual Behavior. *Science* 143:212-218.

## V

## HOMOSEXUALITY

Homosexual practices are among the most ancient manifestations of human sexuality. Abundant evidence for this fact is found in both the religious and the secular literature of the oldest civilizations and in their graphic art as well. Social attitudes in different cultures apparently varied then, as now, from strong condemnation of all homosexual practices to tolerance or permissiveness for some. From the beginning of the Christian era in western Europe ecclesiastical and secular law sought to prevent, control, and eradicate homosexuality by means of severe moral and legal condemnation. By the late nineteenth century, legal penalties were less severe as a result of humanitarian reform, but moral and religious censure had not diminished. That the scientific study of homosexuality began in this period of western European history, and in this cultural climate, is a tribute to the scientific and humanitarian interests of physicians, particularly such pioneers as Richard von Krafft-Ebing,

Sigmund Freud, Havelock Ellis, and Magnus Hirschfeld. Their theories were largely concerned with etiology and were based on clinical observations and influenced by developments in the biological sciences, including evolutionary theory. Although by this time cross-cultural observations of homosexuality were also accumulating, together with the rudiments of a general theory, they had little effect on the mainstream of research, which continued to focus on the etiology of homosexuality among individuals of Western culture.

Increased research on homosexuality in Western countries since World War II has been coupled with more open discussion of the subject. Although there is no reliable evidence that the incidence of homosexuality is increasing, some forms of it are more socially visible. In part, this may be accounted for by the publicity given to scientific investigations of homosexuality and to discussions of the controversial social, legal, moral, and mental health issues related to it. Emergent homosexual "communities" in large urban centers and "homophile" organizations openly protesting public policy further contribute to this heightened visibility and increased public awareness. These developments reflect radical changes in the attitudes of the public and of homosexuals, in the social character of homosexuality, and in the volume and focus of scientific research.

Homosexuality (*homo-* derives from the Greek root meaning "same") includes an extraordinary diversity of dyadic relations and of individual mental states and action patterns. The patterns of social organization that develop when homosexuals seek each other out also vary greatly. Because of this diversity, "homosexuality" is an ambiguous term with many meanings. Some investigators, such as Ford and Beach (1951), limit the term to overt sexual relations between individuals of the same sex. For others, notably Kinsey and his associates (1948; 1953), the degree of psychic arousal and the frequency of overt sexual response to individuals of the same or opposite sex determine ratings on a heterosexual-homosexual continuum that ranges from exclusively heterosexual to exclusively homosexual, with several intermediate ratings. Investigators with a clinical perspective (see, for instance, Marmor 1965, p. 4) frequently stress motivational and subjective aspects of erotic preference as criteria for defining persons as homosexual; overt behavior is considered as of secondary importance. In the social perspective of homosexual subcultures, the defining criteria may be shared understanding of sexual preferences for members of one's own sex and participation in social activity

centered on the search for, and interaction with, these individuals (Goffman 1963, pp. 143-144). Thus, it is clear that "Who is homosexual?" and "What is homosexuality?" are very complex questions, clarification of which would be a lasting contribution to social science.

Research on homosexuality is also of major significance for the light it throws on the relation of social structure and cultural norms to patterns of human sexuality; on the origins, development, and essential features of normal and abnormal personality; on masculinity and femininity; and on social deviance and deviant subcultures.

### Etiology and determinants

In those Western societies that define homosexuality as deviant, there are four major theoretical issues concerning the etiology and determinants of persistent or predominant adult homosexuality. (1) Is the human organism psychosexually neutral at birth, so that learning processes determine homosexual object choice in adults, or are there "inherent sexual predispositions" (Diamond 1965, p. 168) which selectively influence the effects of learning? (2) What is the nature and content of the learning processes by which homosexual object choice develops? Is the appropriate developmental model a deviant role or a personality system with intrapersonal traits, motives, and gender identifications incompatible with the social-sexual capabilities and self-other expectancies of adult relations with the opposite sex? Does positive conditioning of sexual responses to persons of the same sex, or negative conditioning to persons of the opposite sex, or a combination of both, account for homosexuality? (3) Are particular periods in the developmental process, such as early childhood or adolescence, critical for homosexual object choice? (4) Are parent-child relationships in the nuclear family crucial in determining whether an individual becomes homosexual, or are peer relationships in childhood and adolescence, and deviant subcultures in adolescence and early adult life, of equal or possibly greater importance? These four issues are highly controversial and cannot be resolved by the research evidence currently available. A brief historical review of some of the major studies will illustrate differences in theoretical assumptions, in methodological approaches, in findings, and in conclusions, and it will indicate important areas for continued research.

**Theories of homosexuality.** Early theorists stressed the importance of biological predispositions as determinants. Krafft-Ebing (1886) theorized that homosexuality was produced by a

dominance of opposite-sex brain centers (female centers in males or vice versa) or that it was a hereditary disease. Freud (1905) postulated constitutional bisexuality as one of the biological predispositions which, together with early learning experiences or fixations, determined homosexual object choice. The major studies of biological predispositions have focused on genetic, chromosomal, and hormonal anomalies or dysfunctions. Comparisons of predominantly homosexual and heterosexual males with respect to chromosomal abnormalities have resulted in negative findings. In several studies, maternal age and birth order of homosexuals in a sibling series were significantly later than in control groups of heterosexual men. The degree of concordance of homosexuality in identical male twins appears to be significantly higher than in fraternal twins (for a review of these studies, see Pare 1965). The evidence is controversial but suggests that genetic mechanisms may be important. Significant progress could be made by replicating the studies of twins, birth order, and maternal age and by an intensified search for genetic determinants. Homosexual object choice in adults is not affected by hormone therapy and differences between the endocrine systems of homosexuals and those of heterosexuals have not been reliably demonstrated (Perloff 1965). As more accurate techniques of measuring hormonal activity develop, continued research may be valuable. A potential contribution of biological factors to human homosexuality is suggested by recent studies of animals, in which the location and functions of sex-specific behavior centers in the brain have been demonstrated (Krafft-Ebing may be vindicated!). Studies at the infrahuman level have also demonstrated that hormones injected prenatally produce effects on adult sexual behavior (for a review of these studies, see Diamond 1965). Thus, specific biological predispositions may be crucial variables, interacting with psychological and cultural variables, especially in producing the highly feminized males and masculinized females often referred to as "inverts."

The experiential determinants emphasized by classical psychoanalytic theory are pathogenic parent-child relations and failure to successfully resolve (unconsciously) the vicissitudes of the three basic phases of libidinal and object-relations development in the critical period of infancy and early childhood. Some revisions of psychoanalytic theory reject constitutional or psychological bisexuality and emphasize ego adaptations to environmental difficulties. With few exceptions, however (see, for instance, Thompson 1947; Sullivan



1953), pathogenic relations with parents in early childhood are assumed to be the crucial determinants, although individual patterns of behavior are subject to modification during adolescence. (For a review of psychoanalytic theories, see Society of Medical Psychoanalysts 1962, chapter 1.)

This theoretical assumption currently relies for empirical support on conflicting and controversial data derived largely from comparing the life histories of adult homosexual and heterosexual males. Freund and Pinkava (1961) carefully tested clearly formulated hypotheses concerning parent-child relations and homosexuality by comparing interview data from male homosexual patients in a psychiatric hospital with similar data from two contrasting matched groups: neurotic heterosexual patients in the same hospital and nonpsychiatric patients. They found no significant differences between these groups with respect to the dominance, control, or decisive influence of either parent in the family. Relations between homosexual sons and their fathers were more "disturbed," a finding tentatively interpreted by the authors as secondary to, rather than a cause of, homosexuality, since the disturbance appeared to be mainly on the son's side of the relationship. Reports of greater conflict between parents of homosexuals signified a lack of tolerance by their sons for marital conflict rather than greater incidence of actual conflict. Bieber and his co-workers (see Society of Medical Psychoanalysts 1962) compared questionnaire data provided by psychoanalysts concerning their homosexual and heterosexual patients. Some differences between patient-mother-father, mother-son, and father-son relationships in the contrasting groups were statistically significant. The authors concluded that the chances of becoming homosexual or of developing severe homosexual problems appear to be high for any son who is exposed to the combination of a "close-binding-intimate" mother who dominates and minimizes her husband, and a father who is detached or hostile (p. 172). Schofield (1965) compared interview data obtained from six matched groups: (1) homosexual and pedophilic prisoners convicted of homosexual offenses, (2) homosexual and heterosexual patients who were in psychiatric treatment, (3) homosexuals and heterosexuals who were neither in treatment nor in prison. The proportion of individuals from disturbed home backgrounds was similar in each matched pair of groups; it was lowest in the non-treatment-prison groups and highest in the patient groups. The proportion of individuals who reported possessive, overprotective, or dominant mothers

and poor or nonexistent relations with fathers in childhood was larger in the combined homosexual groups (pedophiliacs excluded) than in the combined heterosexual groups (p. 105).

The evidence from these and many similar studies does not support the assumption that pathological parent-child relations are either necessary or sufficient antecedents or determinants of adult homosexuality. The evidence does indicate, however, that some forms of familial pathology appear to be associated with increased vulnerability of some individuals to homosexual development, and it suggests that psychopathology is more frequently associated with homosexuality in these individuals. Longitudinal studies could now be designed for the testing of hypotheses specific to the differential effects of variable combinations of parent-child relations on the development of adult psychosexual patterns. The potential contribution of such studies is very high.

Empirical evidence that determinants in adolescence are critical for adult homosexuality is fragmentary. The theoretical rationale for this assumption is that one of the major developmental tasks of adolescence is the achievement of a stable psychosexual identity. An extensive body of clinical-theoretical literature emphasizes the adolescent period as critical for the resolution of intrapsychic psychosexual conflicts originating in infancy or early childhood (Society of Medical Psychoanalysts 1962). Sullivan (1953) stressed very close and intense preadolescent friendships with peers of the same sex as one of the essential conditions for heterosexual development and the separation of lust and intimacy "dynamisms" as one precondition of adult homosexuality.

Assuming that the human capacity for homosexual as well as heterosexual response is a basic part of the mammalian heritage, Kinsey and his associates (1948) suggested that cultural pressures and social conditioning may determine the final object choice. That adolescence may be a critical period for males is indicated by (1) the existence of a high positive relation between the early onset of adolescence and frequency of homosexual activity during adolescence and later life (*ibid.*, p. 630) and (2) occurrence of the peak of sexual activity between ages 16 and 20. Among females, however, no consistent relation was found between homosexual activity and early onset of adolescence (Kinsey et al. 1953, p. 462), and the peak of female sexual activity occurs in the late twenties (*ibid.*, p. 759).

Group norms and peer relationships in adolescence may determine the relative frequencies of

heterosexual and homosexual patterns in adults by affecting their self concepts, sex role expectations and performances, and notions of what constitute the boundaries of permissible behavior. Reiss (1961) found that members of a group of male juvenile delinquents who engaged in sexual acts with adult homosexuals did not define themselves as homosexuals and did not continue homosexual activity as adults. Group norms defined the limits of permissible sexual activity, rules of payment, frequency of contact, and derogatory attitudes toward homosexuality. Conversely, late adolescents or young adults who participate in homosexual cliques or informal homosexual groups may progressively redefine homosexual behavior in positive terms, defining themselves as homosexuals in the process, and thus beginning a homosexual "career" (Hooker 1965a, pp. 90, 101).

Adult homosexual roles may be formed by a continuous process of social-sexual learning, from early childhood to adolescent and early adult life. Self-other interactions, including overt sexual activities, in deviant socially marginal pairs or groups of peers of the same sex, may be critical determinants of motives, self concepts, sex role expectations and abilities, and performances that are congruent with deviant and homosexual roles and incongruent with heterosexual roles. At critical choice points in the developmental process or continuously from early childhood, situational and social-structural variables may produce effects influencing adult patterns—for instance, the feminizing effects of older sisters on younger brothers, situationally determined isolation from peers of the same sex in childhood and adolescence, or isolation from peers of the opposite sex in adolescence.

From the limited evidence currently available, it is clear that the diverse forms of adult homosexuality are produced by many combinations of variables, including biological, cultural, psychodynamic, structural, and situational. No single class of determinants, whether psychodynamic, cultural, or biological, accounts for all or even one of these diverse forms. The relative importance of each kind of determinant appears to vary greatly from one individual to another. It should be noted, however, that the classification of types of homosexuality with respect to patterns of determinants and characteristic adult features is an important problem for research. It is also clear that many developmental models are required to account for the diversity of developmental sequences leading to confirmed patterns of adult homosexuality. The continued search for determinants may be more

productive if pursued in the perspective of development over time, rather than in the traditional perspective of origins or causes.

### Psychodynamics of adult homosexuality

Psychoanalytic theories assume that homosexuality is a symbolic expression of unconscious psychodynamic mechanisms and that in any given case these mechanisms and symbolizations are complex and numerous. Thus, according to Freud (1905), male homosexuality may be a narcissistic search for a love object symbolizing the self; an avoidance and derogation of females induced by unconscious castration anxiety and fear of them, coupled with primary narcissism focused on the male genitals; and an unconsciously incestuous and therefore forbidden attraction to women associated with unconscious castration anxiety and fear of punishment by males. Stronger identification with the mother is symbolically expressed in "feminine" sexual submission to males; that with the father, in "masculine" domination of males. Freud postulated similar complex symbolic expressions of unconscious psychodynamics for female homosexuality; the primary difference between male and female homosexuality, he thought, related to the nature of castration anxiety. However, there is a paucity of theoretical assumptions that differentiate male and female homosexuality with respect to personality dynamics and its correlates (see Romm 1965). Ambivalence, chronic anxiety, intense hostility toward the partner, and longing for love characterize female homosexual relationships, according to Wilbur (1965). Thompson (1947) considered homosexuality in both sexes as a symptom of diverse character problems covertly expressed in homosexual relations (among these problems were fear of the opposite sex, need to defy authority, fear of adult responsibility, and destructiveness of self or others). These theories involve complex patterns of variables that are difficult to state as clear hypotheses. Consequently, the design of an appropriate methodology for testing them presents problems. Research studies have therefore focused on selected aspects of the theories or variables for which clinical impression or theoretical rationale provides justification.

**Masculinity and femininity.** A major focus of research studies on male homosexuality is the complex issue of psychological masculinity-femininity and its relation to gender identification. Terman and Miles (1936), using masculinity-femininity test scores and interview data, compared "passive" male homosexuals (defined by "female" roles in sexual acts) with "actives" (de-



finer by "male" roles) and then compared both of these with (presumably) selected groups of heterosexual men and women. The femininity scores of the passives were higher than those of any other male group tested. The masculinity scores of the actives were higher than those of the passives and of a group of regular-army men. These results suggested that individuals in both homosexual groups represented deviations in masculine and feminine identifications, interests, activities, and feelings. Interview data on the passives supported the masculinity-femininity test results. Similar results were obtained on a small sample of active female homosexuals.

In subsequent studies of masculinity-femininity, data on sexual practices and preferences are rarely reported, but the evidence from research on other dimensions of male homosexuality indicates that samples of homosexuals usually include many individuals whose preferences and practices are highly variable, as well as those representing the two extremes of active and passive (Society of Medical Psychoanalysts 1962; Hooker 1965*b*, Schofield 1960). The distribution of the scores of homosexuals and control groups of heterosexuals on the masculinity-femininity scale of the Minnesota Multiphasic Personality Inventory (MMPI) indicates that homosexuals are significantly more "feminine" in their responses than heterosexuals. Thus, Doidge and Holtzman (1960, p. 11) found that a group of homosexual men in the Air Force who had been investigated for homosexual offenses scored significantly higher (that is, were more "feminine") than three control groups of predominantly or exclusively heterosexual men. Braaten and Darling (1965, p. 298) found that, while overt homosexual males (including exclusives and bisexuals) and covert homosexuals who were patients in a college clinic did not differ significantly on masculinity-femininity scores, as a total group (overts plus coverts) they scored significantly higher than a control group of equally disturbed, nonhomosexual patients. Although a group of homosexual males who were not in treatment and were apparently functioning effectively in the community scored significantly higher than a control group of heterosexual males on the masculinity-femininity scale of the MMPI and the femininity scale of the California Personality Inventory (CPI), significant differences were not found in scores on various projective measures designed to assess unconscious gender identification (unpublished research by the author). Feminine identification and confused or unstable gender identification in male homosexuals have

also been assessed by the use of content indices in the Rorschach Test and the Thematic Apperception Test (TAT). The interstudy reliability of these indices in differentiating homosexual and heterosexual groups is low (Hooker 1965*a*; Lindzey 1965). Chang and Block (1960), using a list of adjectives, derived identification scores from descriptions by male homosexual and heterosexual research subjects of the ideal self, the mother, the father, and their actual selves. The results supported the hypothesis that homosexual males were more strongly identified with the mother and did not identify with the father.

Conclusions from these findings must be drawn with caution. Psychological masculinity and femininity are not unitary dimensions of personality but composites of different clusters of interests, attitudes, and feelings. The test measures are also multivariate in character. Thus, high "femininity" scores may indicate ego sensitivity and denial of interest in culturally masculine occupations (Braaten & Darling 1965). Research findings currently available do not provide even roughly sketched outlines of the complex patterns of psychological masculinity-femininity and gender identification in male or female homosexuals. If current measures of these variables are to remain in use, the multivariate character of what is being measured must be taken into account. Also, new and more appropriate methods must be devised for assessing different clusters or profiles of the conscious, unconscious, cultural, and biological components of masculinity, femininity, and gender identification. The selection of samples of research subjects must make allowance for the broad range and diversity of masculine and feminine patterns and the variable combinations of these.

**Other personality dimensions.** The evidence supporting assumptions related to other psychodynamics or personality correlates is equally inconclusive. Bieber and his co-workers (Society of Medical Psychoanalysts 1962) concluded that their findings provide convincing support for the assumption that homosexuality is a specific form of sexual psychopathology: fear of the opposite sex or of heterosexual expression. But none of the evidence, such as fear of injury to the genitals, fear and aversion to female genitalia, or anxiety associated with heterosexual behavior, was specific to the homosexual group. The evidence cited for the conclusion that other psychopathologic processes in personality are secondary but invariably occur in homosexuals is subject to the same limitations. Moreover, the research subjects were patients in psychoanalytic treatment, so that the relation be-

tween psychopathology and homosexuality was confounded. Braaten and Darling (1965) found no evidence to support assumptions that homosexuals, as rated by psychiatric interviewers, were more psychopathic, paranoid, or schizoid than a control group of heterosexuals and found no differences in psychopathology as measured by the MMPI. Psychiatric ratings on narcissism and masochism were higher for homosexuals but not specific to them. The significantly higher scores of homosexuals on all clinical scales of the MMPI were interpreted by Doidge and Holtzman (1960) as evidence that homosexuality is a symptom of severe personality disturbance. In a number of studies of male homosexuals who were not in treatment or in disciplinary settings, the results obtained by the use of the MMPI, TAT, Rorschach, and other psychological measures did not justify the conclusion that homosexuality is necessarily and invariably a concomitant or symptom of psychopathology. In many individuals no evidence of psychopathology was found (Hooker 1957; 1965a). Nevertheless, the assumption that a modal pattern of personality structure and psychodynamics characterizes male homosexuals is explicitly stated or is implicit in the methodology used in many research studies. Having been unable to establish any differences between a group of homosexual males and a control group of heterosexuals, either in degree of self-acceptance or in ideal-self descriptions, Chang and Block commented that since "homosexuals and normals are quite obviously rather different sorts of people," their data must be regarded as questionable (1960, p. 309). Nevertheless, the only obvious difference between homosexuals and heterosexuals is in psychosexual object choice. All experienced clinicians and research workers report that the personality differences among individual homosexuals are far more apparent than the similarities. Investigators who include the data on individual differences in their studies have found a great diversity of personality patterns. A singular exception is the finding of Cattell and Morony (1962) that low ego strength, high extroversion, low superego, and a radical social outlook characterize the personality profiles of male homosexuals. They conclude that it is legitimate to speak of *the* homosexual-personality profile.

The inconclusiveness of research on homosexuality is a function of many variables, including differences in investigators, in samples of research subjects and situational settings, in methodology, and in concepts of homosexuality. Lindzey (1965) has demonstrated the significant influence of many of these variables in a study designed to

cross-validate objective indices of homosexuality in the TAT and to test the predictive efficiency of clinical methods of diagnosing homosexuality compared with actuarial ones. The limitations of psychological tests in the diagnosis of homosexuality have been documented in many studies. Inferences from these measures and generalizations about the personality characteristics of homosexuals must be made with caution and within the limits set by the characteristics of the research sample, the setting, and the methodology.

With few exceptions, studies of the personalities of homosexuals have been stated in terms of dichotomous attributes—psychopathology versus normality, specific sexual psychopathology versus pervasive personality disorder, feminine versus confused identification, feminine versus masculine identification, and other pairs of mutually exclusive variables. Thus, in effect, homosexuality has been treated as a clinical entity. But contemporary personality theories are richly suggestive of dimensions or subsystems of personality that are potentially relevant to psychosexual patterns, and until now, these theories have scarcely been tapped as sources of hypotheses. Some of the effects of a changing climate of public attitudes toward homosexuality and of an increasing freedom to conduct scientific studies of human sexuality in all of its complex psychological, social, and biological aspects are manifested in the small but growing number of behavioral scientists whose interests and competence in personality research are focused on homosexuality. With the cooperation of homosexuals in the "homophile" and larger communities, it is no longer necessary to depend on subjects in clinic or prison settings. With the use of more rigorous criteria in selecting homogeneous samples of subjects and of more appropriate research designs, including cluster and profile analyses, it may be possible to differentiate personality patterns among homosexuals. Also, progress may be achieved in resolving the long-debated issue as to whether some homosexuals differ from their heterosexual counterparts only in sexual object choice or are invariably different in some personality dimensions.

### Role differentiation and typology

The perspectives of the two-sexed heterosexual society have so dominated the study of homosexuality that typologies of homosexual and related roles have been developed only to a very limited extent. Thus, Ferenczi (1914) distinguished between male inverts who have pervasive feminine identifications and assume a passive sexual role



and males who assume an active sexual role and are apparently masculine in every respect except sexual object choice. Many theoreticians and empirical investigators (see, for instance, Terman & Miles 1936), while emphasizing the active-passive dichotomy in sex roles and the corresponding masculine-feminine dichotomy in gender identifications, have noted that the patterns of many male homosexuals represent variable combinations and intermediate grades. Recent studies (Schofield 1960, Society of Medical Psychoanalysts 1962, Hooker 1965b) have shown that the terms "activity" and "passivity" may be useful in referring to psychological states but are inaccurate and misleading when applied to sex roles and should therefore be discarded. The terms "insertee" and "insertor" (Society of Medical Psychoanalysts 1962) are more accurate and operationally useful equivalents. Studies of the sex-role preferences and practices of male homosexuals indicate that a large proportion of them habitually perform either role, according to their own or their partners' momentary preference. For the majority, consciousness of male identity and of masculinity or femininity appears to show no clear relation to sex roles. Sustained dyadic relations and the shared perspective of male homosexual subcultures appear to increase the flexibility of social-sexual roles and to create the social conditions for the emergence of new solutions to the problems of role differentiation in a one-sex world (Hooker 1965b).

Studies of male homosexuality in correctional institutions (Huffman 1960; Sykes 1958) suggest that homosexual social-sexual roles are clearly defined by inmate codes. According to these codes, the masculinity of the individual who assumes the insertor role in a homosexual act is not diminished and may even be increased. Within the institution, the role carries no implication that the individual is homosexual. Emotional involvement with a partner appears to be rare. Gratification of needs for power, status, and sexual release is presumably legitimate motivation for assuming the role. Individuals who assume insertee roles occupy social positions of inferior status to which stereotyped attitudes of femininity and homosexuality are ascribed. These traditions apply only to institutions in which known homosexuals are not segregated from the rest of the prison population.

As noted above, the literature of female homosexuality is fragmentary in comparison with that of male homosexuality. Studies of role differentiation have been conducted largely in correctional settings for adolescent girls or prisons for adult

women. In prisons for women a male-female dichotomy in social-sexual roles develops and is sustained with little flexibility or variation (Ward & Kassebaum 1965). The masculine, or "butch," role is characterized by emulation of the male in appearance, mannerism, and aggressiveness in social interaction as well as in sexual relations. The feminine role is identical in the prison and in the larger society, except for the sexual partner. The woman who assumes the masculine role becomes emotionally involved with her feminine role partner and defines herself as homosexual more frequently, and receives sexual gratification less frequently, than her counterpart in the male prison. Whether the clear-cut male-female role differentiation in prison settings is also characteristic of female homosexual patterns in the larger community is unknown. Despite the differences, the similarities of homosexual role differentiation in men's and women's prisons are striking. It appears that individuals in both settings who assume opposite-sex roles are more likely to have been overt homosexuals before entering prison and to continue homosexuality on being released. In studies of correctional settings for adolescent girls, it has been found that the roles assumed in the course of homosexual relationships represent variations of all family roles.

The empirical data obtained in the restricted environments of correctional institutions for men and women support a dichotomous typology of homosexual roles in which the male and female social-sexual roles in the larger community are reflected. However, male homosexual role patterns in other cultural settings appear to show extensive variations for which a taxonomy, or system of classification, is urgently needed. Studies of female homosexual role patterns in nonprison settings—several were in progress in 1964—can provide the necessary data for a comparison of variability within and between the sexes. The development of taxonomies of male and female homosexual patterns is a theoretical problem of paramount importance. To be meaningful, such taxonomies must include variations in many dimensions of personality, in sexual and social behavior, and in the social settings or subcultures in which many homosexuals participate.

**Demographic and cultural variation.** In undertaking empirical research on the incidence or demographic distribution of homosexuality in urban industrialized societies, investigators are confronted with two major questions: 1. What or what is to be counted as homosexual? and 2. How can representative samples be obtained in

of the sexual status of individuals whose behavior is socially, morally, and legally stigmatized." In view of the magnitude of these problems, it is not surprising that few have even attempted to solve them. The most systematic research studies of the incidence and social correlates of homosexuality are those of Kinsey and his associates (1948, 1953), who rejected the counting of homosexuals as an inappropriate and impossible task. Accumulative estimates of incidence were based on the number of individuals brought to orgasm by physical contact with others of the same sex during specified periods. They attempted to collect interview data from samples of equal size, varying in demographic characteristics. The samples (6,300 white males, 5,940 white females) were neither representative of the total population nor random around them.

The Kinsey studies the estimated incidence was reduced by the following: (1) 4 per cent had at least one overt homosexual experience between the ages of adolescence and old age, (2) 10 per cent were predominantly homosexual for at least three years between the ages of 16 and 75, (3) 4 per cent were solely homosexual after the onset of adolescence. For white females, the estimated incidence was much lower: (1) 13 per cent had at least one overt homosexual experience between adolescence and old age, (2) 28 per cent had covert or overt experience, (3) a half to a third as many females were primarily or exclusively homosexual in any age period.

The major findings on the social correlates of homosexual activity for males were that the highest incidence occurred among unskilled and semiskilled workers and the lower white-collar class, among those with high school education and of other religious, and among the less devout or inactive church members. The lowest incidence occurred among the professional and upper white-collar classes and in the college-educated group. However, the incidence among skilled workers was closer to that among the professional and upper white-collar classes than to that among the lower occupational groups. Those with grade school education showed a level of homosexual activity intermediate between that of high school and that of college-educated groups. Homosexual activity in females increased with education so that the incidence was highest among those with college or postgraduate education. For males, the opposite was true. For both males and females the highest incidence occurred among the less devout or least active in church affiliation.

Although the assumptions, methods, and conclu-

sions of these studies are highly controversial, their findings are most frequently used as evidence for a high incidence of homosexuality in the United States. They are also based on the basis of rather limited demographic evidence to support the assumption that cultural differences between social groups determine the unequal distribution of homosexual activity in different segments of society.

This assumption also appears to be supported by Schofield's study in England, 1965. A group of 127 men who identified themselves as homosexuals to the investigator and were also rated by him as predominantly or exclusively so were interviewed concerning their backgrounds and current social and sexual patterns. The largest proportion of men were in lower occupational levels and the smallest in the professional group, a distribution of men among occupational levels that is similar to the total population. However, the number of men in nonmanual and artistic occupations and in the professions exceeded chance expectations and those employed in a technical capacity and manual workers were underrepresented. In educational level attained, the sample was similar to the total population in that the largest number had the statutory minimum of education. However, the number of those with extensive education and high social status was higher than would be expected in a random sample of the total male population. The majority of these men lived in London to which a large proportion of them had migrated from other urban centers. Active church affiliation was only slightly lower than that of the general population. In a continuation of this study, fifty men were selected who had been neither convicted of homosexual offenses nor involved in therapy of any kind (Schofield 1965). They were compared with an equal number of men in three groups: homosexuals in prison, homosexuals in therapy, and pedophiles convicted of offenses against males. On Kinsey ratings, the prison group was predominantly or exclusively homosexual, the patient group included a small number who were equally homosexual and heterosexual, or predominantly heterosexual, a quarter of the convicted homosexuals and 60 per cent of the pedophiles were either equally homosexual and heterosexual or predominantly heterosexual. In educational achievement, occupational status, earned income, social class and upward social mobility, the men who were neither in prison nor in therapy ranked highest and the prison groups lowest. The four groups were differentiated not only in their demographic characteristics but also in their patterns of social and



sexual relationships. Thus, the men who had not sought treatment or been arrested were better integrated into the larger community both at work and in their leisure time. They also had larger circles of homosexual friends and more long-standing relationships with other men. Only in this group were there substantial numbers of men who maintained a household with another homosexual. They were far less promiscuous than the patient or prison homosexual groups, although the frequency of sexual experience was equally high.

It is apparent that comparison of the findings in this British study with those in the Kinsey report has limited meaning. However, some similarities and differences are of interest. In both studies, homosexuals were found in all social classes, educational levels, and occupations. A larger proportion of homosexuals (or of homosexual activity) in all studies took place in urban settings and in lower occupational levels and social classes. In educational achievement, however, the proportion was greater in the middle level (high school) in America and in the lower level (secondary modern school) in England. That in both countries the distributions were skewed toward the lower end of the occupational and social class range may well be accounted for by the fact that individuals in higher levels would not volunteer to be interviewed. The same limitation is relevant to the Kinsey finding that the incidence of homosexual activity is lowest among the most devout religious groups. Moreover, in the English studies the incidence of active religious affiliation did not appear to be significantly lower among homosexual males.

In short, whether homosexuals or homosexual activities among males or females are more heavily concentrated in different social strata remains an open question. Whether the social patterns or incidence of those who *define* themselves as homosexual are different in varied social levels is of equal or possibly even greater importance.

**Ethnographic accounts.** In a survey of studies of sexual behavior in 190 societies, Ford and Beach (1951) found that ethnographic accounts of homosexual activities were available for 76. In 49 of these some forms were condoned, regarded as normal or socially acceptable, and even encouraged for at least some classes of individuals. In some cultures, such as those of the Siberian Chukchee and the Aleutian Koniaga, an institutionalized role of berdache, or shaman, is provided for adult male homosexuals. These men adopt feminine dress, activities, mannerisms, become "wives" of other men, and assume the "female" role in anal inter-

course. Their social status may be high. Male homosexuality in some societies, for instance among the Keraki and Kiwai of New Guinea, is an institutionalized feature of puberty rites. Thus, all males must engage in homosexual practices, either as initiates in the insertee role in anal intercourse or as married or unmarried males who perform the inserter role. Such practices are believed to be essential for male growth and strength. It has also been reported that all men and boys among the Siwans of northeast Africa engage in homosexual practices; married and unmarried males have homosexual and heterosexual liaisons. In 28 of these societies, adult homosexuality was reported to be rare, absent, or only practiced in secret. In all societies (except that of the Sirione in Bolivia) in which homosexuality was said to be rare, specific social pressures were directed against it. Condemnation and penalties ranged from ridicule to threat of death. The unreliability of the estimates of incidence in these societies was apparent. Information concerning female homosexuality was reported for only 17 societies. It appeared highly probable that females were less likely than males to engage in homosexual practices.

In view of the small number of societies for which data are currently available and the questionable reliability and meaning of the data in the many studies that report sexual behavior while ignoring its context of culture patterns and social structures, it is apparent that descriptive or explanatory generalizations concerning demographic and cultural variations in homosexuality are precariously based. Whether a majority of societies consider some form of homosexuality normal and socially acceptable is an important issue for continued cross-cultural research. Clearly, some societies not only consider adolescent or childhood homosexuality normal (some evidence suggests an increasing trend in this direction in the United States) but also regard homosexual relations between adult and adolescent males as important, normal, and socially approved alternatives to heterosexual intercourse (Davenport 1965).

**Subcultures and social organizations.** Homosexual communities and subcultures in modern urban centers are collective reactions to legal pressures and social stigma. Whether they are largely created by societal reactions to this form of sexual expression (Becker 1963) or, on the contrary, would develop under any sociolegal conditions is debatable (Schur 1965, p. 86). A central thesis in contemporary theories (Lemert 1951; Becker 1963; Goffman 1963) is that deviant as well as conform-

ing behavior is learned in interaction with others and that it cannot be understood without reference to the societal reactions it invokes. Male homosexual collectivities develop on a sizable scale only in modern industrialized societies, which define this form of sexual expression as criminal, pathological, or immoral. In adjusting to this stigmatized status, a large number of homosexuals in social interaction with others who share similar inclinations make it the basis of their social identity and way of life. A larger majority cope with the problems of homosexuality in other ways, whether or not they enter some sector of a homosexual community.

Homosexual communities in large cities are made up of constantly changing aggregates of persons who are loosely linked by friendship and sexual interests in an extended and overlapping series of networks. Some network clusters form tightly knit cliques of friends and homosexually "married" pairs, while in others, informal groups or social organizations develop. The structure of the communication, friendship, and sexual network among members of the community is complex. Community gathering places are centers from which information is transmitted concerning social occasions for homosexuals and the attitudes and organized activities of law-enforcement agencies concerning homosexuals. In some sectors of these networks, bars, taverns, or private clubs are informally institutionalized as homosexual territory, although under surveillance of the police, who sometimes exact tribute for their connivance. These settings provide opportunities for the initiation of sexual contact and for an in-group social life. Some are primarily market places for making sexual contacts, while others are predominantly centers of social activity, although arrangements for sexual activity may occur. A standardized and essential feature of interaction in bars, baths, streets, and parks is the expectation that sex can be had without obligation or commitment. Sexuality is separated from affectional and social life and is characterized by promiscuity, instrumentality, and anonymity. In other sectors of the homosexual community, sexuality is integrated in the affectional, personal, and social patterns of individuals who establish relatively stable and long-lasting relationships.

Many subcultures develop in large homosexual communities, with a distinctive vernacular, distinctive modes of speech and dress, distinctive values and norms regulating sexual and other forms of conduct, and distinctive patterns of interaction

with homosexuals and nonhomosexuals. Some individuals may be totally immersed in a homosexual subculture of the community. A wide variety of occupations is created by the demand for services to homosexuals. Male prostitutes, many of whom are homosexuals, constitute one occupation (Reiss 1961; Schur 1965). Entrepreneurs who are homosexual may develop an occupation in arranging social entertainment and large social gatherings for members of a particular homosexual group. Thus, at work and in leisure time, in personal and social life, contact with the larger society may be infrequent. For the majority, however, behavioral, temporal, and spatial separation of the homosexual and heterosexual worlds is required to manage the strategies of passing as heterosexuals at work and in social contact with heterosexuals. In fact, there are many different methods, some of them quite complex, both of participating in the homosexual community without leaving the larger society and of ignoring the larger society altogether. Related to these patterns of integration and separation are the different social levels of homosexual groups.

Female homosexual, or lesbian, communities apparently develop on a smaller scale, with informal groups, cliques, and special gathering places. But a formal organization of lesbians and an official publication with national circulation indicate that the collective aspects of female homosexuality have some importance. No empirical studies of these aspects are currently available. Although homophile organizations, whether male or female, constitute and represent a very small minority of the total homosexual population, they achieve social significance by the role they assume in openly protesting the status assigned to homosexuals by the larger society.

Empirical studies of, and theoretical inquiry into, homosexual subcultures are very limited (Leznoff & Westley 1956; Hooker 1965a; 1965b). They suggest that commitment to a homosexual identity and career and the stability of that commitment and the career pattern that follows may be determined by entry into the homosexual community and by the ensuing identity transformation and socialization processes. However, the relation between personality and homosexual subculture variables in determining the commitment to, and patterns of, adult homosexuality is complex. For many the stability of the commitment appears to be a function of the interaction of both sets of variables. Probably a majority of those who are committed to an exclusive homosexual pattern have defined themselves as homosexual before they



enter the community. In major urban centers, however, it appears unlikely that adult homosexuals are ignorant of, or indifferent to, the community.

EVELYN HOOKER

[See also PERSONALITY; and the biographies of ELLIS; FREUD.]

#### BIBLIOGRAPHY

- Reviews of research studies may be found in Marmor 1965; Ruitenbeek 1963; Schofield 1960, 1965. Kinsey et al. 1948; 1953 include surveys of earlier work on homosexuality.
- BECKER, HOWARD S. 1963 *Outsiders: Studies in the Sociology of Deviance*. New York: Free Press.
- BRAATEN, LEIF J.; and DARLING, C. DOUGLAS 1965 Overt and Covert Homosexual Problems Among Male College Students. *Genetic Psychology Monographs* 71: 269-310.
- CATTELL, RAYMOND B.; and MORONY, JOHN H. 1962 The Use of the 16 PF in Distinguishing Homosexuals, Normals, and General Criminals. *Journal of Consulting Psychology* 26:531-540.
- CHANG, JUDY; and BLOCK, JACK 1960 A Study of Identification in Male Homosexuals. *Journal of Consulting Psychology* 24:307-310.
- DAVENPORT, WILLIAM 1965 Sexual Patterns and Their Regulation in a Society of the Southwest Pacific. Pages 164-207 in Conference on Sex and Behavior, Berkeley, California, *Sex and Behavior*. Edited by Frank A. Beach. New York: Wiley.
- DIAMOND, MILTON 1965 A Critical Evaluation of the Ontogeny of Human Sexual Behavior. *Quarterly Review of Biology* 40:147-175.
- DOIDGE, WILLIAM T.; and HOLTZMAN, WAYNE H. 1960 Implications of Homosexuality Among Air Force Trainees. *Journal of Consulting Psychology* 24:9-13.
- FERENCZI, SÁNDOR (1914) 1963 The Nosology of Male Homosexuality. Pages 3-16 in Hendrik M. Ruitenbeek (editor), *The Problem of Homosexuality in Modern Society*. New York: Dutton. → First published in German.
- FORD, CLELLAN S.; and BEACH, FRANK A. 1951 *Patterns of Sexual Behavior*. New York: Harper.
- FREUD, SIGMUND (1905) 1953 Three Essays on the Theory of Sexuality. Volume 7, pages 123-245 in *The Standard Edition of the Complete Psychological Works of Sigmund Freud*. London: Hogarth; New York: Macmillan. → First published in German.
- FREUND, K.; and PRINEAVA, V. 1961 Homosexuality in Man and Its Association With Parental Relationships. *Review of Czechoslovak Medicine* 7:32-40.
- GOFFMAN, ERVING 1963 *Stigma: Notes on the Management of Spoiled Identity*. Englewood Cliffs, N.J.: Prentice-Hall.
- HOOKE, EVELYN (1957) 1963 The Adjustment of the Male Overt Homosexual. Pages 141-161 in Hendrik M. Ruitenbeek (editor), *The Problem of Homosexuality in Modern Society*. New York: Dutton.
- HOOKE, EVELYN 1965a Male Homosexuals and Their "Worlds." Pages 83-107 in Judd Marmor (editor), *Sexual Inversion: The Multiple Roots of Homosexuality*. New York: Basic Books.
- HOOKE, EVELYN 1965b An Empirical Study of Some Relations Between Sexual Patterns and Gender Identity in Male Homosexuals. Pages 24-52 in John Money (editor), *Sex Research: New Developments*. New York: Holt.
- HUFFMAN, ARTHUR V. 1960 Sex Deviation in a Prison Community. *Journal of Social Therapy* 6:170-181.
- INDIANA UNIVERSITY, INSTITUTE FOR SEX RESEARCH 1965 *Sex Offenders: An Analysis of Types*, by Paul H. Gebhard, J. H. Gagnon, W. B. Pomeroy, and C. V. Christenson. New York: Harper.
- KINSEY, ALFRED C. et al. 1948 *Sexual Behavior in the Human Male*. Philadelphia: Saunders.
- KINSEY, ALFRED C. et al. 1953 *Sexual Behavior in the Human Female*. Philadelphia: Saunders.
- KRAFFT-EBING, RICHARD VON (1886) 1965 *Psychopathia sexualis*. New York: Putnam. → First unexpurgated edition in English; original in Latin.
- LEMERT, EDWIN M. 1951 *Social Pathology: A Systematic Approach to the Theory of Sociopathic Behavior*. New York: McGraw-Hill.
- LEZNOFF, MAURICE; and WESTLEY, WILLIAM A. (1956) 1963 *The Homosexual Community*. Pages 162-174 in Hendrik M. Ruitenbeek (editor), *The Problem of Homosexuality in Modern Society*. New York: Dutton.
- LINDZEY, GARDNER 1965 Seer Versus Sign. *Journal of Experimental Research in Personality* 1:17-27.
- MARMOR, JUDD (editor) 1965 *Sexual Inversion: The Multiple Roots of Homosexuality*. New York: Basic Books.
- OPFLER, MARVIN K. 1965 Anthropological and Cross-cultural Aspects of Homosexuality. Pages 108-123 in Judd Marmor (editor), *Sexual Inversion: The Multiple Roots of Homosexuality*. New York: Basic Books.
- PARE, C. M. B. 1965 Etiology of Homosexuality: Genetic and Chromosomal Aspects. Pages 70-80 in Judd Marmor (editor), *Sexual Inversion: The Multiple Roots of Homosexuality*. New York: Basic Books.
- PERLOFF, WILLIAM H. 1965 Hormones and Homosexuality. Pages 44-69 in Judd Marmor (editor), *Sexual Inversion: The Multiple Roots of Homosexuality*. New York: Basic Books.
- REISS, ALBERT J. JR. (1961) 1963 The Social Integration of Queers and Peers. Pages 249-278 in Hendrik M. Ruitenbeek (editor), *The Problem of Homosexuality in Modern Society*. New York: Dutton.
- ROMM, MAY E. 1965 Sexuality and Homosexuality in Women. Pages 282-301 in Judd Marmor (editor), *Sexual Inversion: The Multiple Roots of Homosexuality*. New York: Basic Books.
- RUITENBEEK, HENDRIK M. (editor) 1963 *The Problem of Homosexuality in Modern Society*. New York: Dutton.
- [SCHOFIELD, MICHAEL] 1960 *Minority: A Report on the Life of the Male Homosexual in Great Britain*, by Gordon Westwood [pseud.]. London: Longmans.
- SCHOFIELD, MICHAEL 1965 *Sociological Aspects of Homosexuality*. Boston: Little.
- SCHUR, EDWIN M. 1965 *Crimes Without Victims*. Englewood Cliffs, N.J.: Prentice-Hall. → See especially pages 1-11 and 67-119.
- SOCIETY OF MEDICAL PSYCHOANALYSTS 1962 *Homosexuality: A Psychoanalytic Study*, by Irving Bieber et al. New York: Basic Books.
- SULLIVAN, HARRY STACK 1953 *The Interpersonal Theory of Psychiatry*. Edited by Helen Swick Perry and Mary Ladd Gawel. New York: Norton.
- SYKES, GRESHAM M. 1958 *The Society of Captives: A Study of a Maximum Security Prison*. Princeton Univ. Press.

- TERMAN, LEWIS M.; and MILES, CATHERINE C. 1936 *Sex and Personality: Studies in Masculinity and Femininity*. New York: McGraw-Hill.
- THOMPSON, CLARA (1947) 1963 Changing Concepts of Homosexuality in Psychoanalysis. Pages 40-51 in Hendrik M. Ruitenbeek (editor), *The Problem of Homosexuality in Modern Society*. New York: Dutton.
- WARD, DAVID A.; and KASSEBAUM, GENE G. 1965 *Women's Prison: Sex and Social Structure*. Chicago: Aldine.
- WILBUR, CORNELIA B. 1965 Clinical Aspects of Female Homosexuality. Pages 268-281 in Judd Marmor (editor), *Sexual Inversion: The Multiple Roots of Homosexuality*. New York: Basic Books.

## SHAMANISM

See RELIGIOUS SPECIALISTS.

## SHANG YANG

Shang Yang (also known as Wei Yang or Kung-sun Yang), who died in 338 B.C., was a Chinese political administrator who left his native state of Wei to enter the service of Duke Hsiao of the state of Ch'in. At the time, the feudal hierarchy dominated by the Chou dynasty had degenerated into a struggle for power among feudal principalities, and Duke Hsiao's consuming concern was to build up Ch'in strength and become a leader among the contending states.

Although it is questionable whether Shang Yang himself wrote *The Book of Lord Shang*, which has been ascribed to him, the conception of a state that is presented in this work is very likely similar to his, and it was therefore surely influential in the reforms he instituted to fulfill Duke Hsiao's ambition. "The central idea of the political theory expounded in *The Book of Lord Shang* is the necessity for creating a strong government; the strength of a government is relative to the weakness of the people and is insured by a strict rule of law" (Duyvendak 1934, p. 16).

Power, for Shang Yang, consisted of a large army and a full granary. He was convinced that the maximum strength of the state could be obtained only through state-wide mobilization, and this in turn required a system of effective laws enforced by severe punishment. He succeeded in instituting laws that applied uniformly throughout the state and equally to all classes of people. The laws were codified and made public, and a judiciary was set up that was independent of the administrative offices of the government. In addition, Shang Yang proposed and implemented a new system of land division, based on private ownership, together with a tithing system of taxation. In accordance with the principle of centralized govern-

ment, military service was owed only to the state, and all private military action was prohibited.

Shang Yang's administrative system was a success. In ten years the state was reputedly rid of lawlessness; it was highly centralized and bureaucratized; and it pursued a victorious policy of foreign war. When the Ch'in state succeeded in subduing all its rivals and reunifying the empire in 221 B.C., the programs initiated by Shang Yang were introduced and extended throughout the empire.

Shang Yang did not long survive the death of Duke Hsiao in 338 B.C. His severe and egalitarian laws had aroused much resentment on the part of the nobility and the privileged classes, and they took the first opportunity to turn against him. His attempts first at flight and then at armed resistance failed, and he was put to death.

*The Book of Lord Shang* remains a valuable record of the social and political philosophy of the ancient Chinese school of law in general and of Shang Yang in particular. The "school of law" stressed the importance of government by law, as opposed to the Confucian ideal of government by "moral virtue." Traditional China, dominated by Confucianism, therefore regarded Shang Yang and the other members of the legalist school with suspicion and abhorrence and condemned Shang Yang for having no moral sense or human feeling. Modern China has re-evaluated him more favorably. The ideas and institutions associated with his name now appear to contain a number of elements considered necessary to the modern administration of law and government. In Shang Yang's approach to politics, however, an informed public played no role, nor were there any safeguards against a ruler who wished to substitute his own whims for the welfare of the state.

Y. P. MEI

[See also CHINESE POLITICAL THOUGHT and the biographies of KAUTILYA; MACHIAVELLI.]

### BIBLIOGRAPHY

- DUYVENDAK, J. J. L. 1934 Shang Yang. Volume 14, pages 15-16 in *Encyclopaedia of the Social Sciences*. New York: Macmillan.
- KUNG-SUN, YANG *The Book of Lord Shang*. Translated by J. J. L. Duyvendak. Univ. of Chicago Press, 1963.

## SHERRINGTON, C. S.

Sir Charles Scott Sherrington (1857-1952) made an overwhelming contribution to the foundations of modern neurophysiology through his ex-



periments and theories on the functioning of the spinal reflexes. Sherrington's importance to the behavioral scientist lies in the fact that he used mainly the methods of behavioral science to pursue his investigations. He studied the behavior of animals after the influence of higher neural centers had been removed surgically by spinal transection. From such behavior he formulated hypotheses about the nature of the nervous system, which he tested by further experiment.

Sherrington was born in London, the son of a country doctor. Although his curriculum at school did not include science, he later enrolled as a medical student at St. Thomas's Hospital in London. Before he completed his clinical studies there in 1885 he worked in the physiology school at Cambridge University and he also spent almost a year with Friedrich Goltz and Carl Anton Ewald in Strasbourg. It was in Goltz's laboratory that he saw the reactions of a dog with a high spinal transection. The investigation of such phenomena was to form his major scientific interest until 1935, the date of his retirement from the Waynflete professorship of physiology at Oxford. Although something was known of reflex action in a general way, Sherrington's thorough investigations transformed the study of reflex action into a tool for the understanding of nervous activity.

His book *The Integrative Action of the Nervous System* (1906) summarizes his most important contributions. The problem that Sherrington posed himself was how the truncated nervous system of the spinal animal produces motor reactions that are coordinated both in space and in time when stimuli are applied. He saw that stimulation of motor pathways alone leads to contortions rather than to actions and that the properties possessed by nerve trunks alone are not sufficient to endow the nervous system with the means of achieving the behavior that he observed. To account for the difference in reaction to different stimuli Sherrington had recourse to the notion of an adequate stimulus, or the idea that the function of a receptor is to make available a particular reflex only to specific physical stimuli. Furthermore, Sherrington was struck by the presence of an after-discharge (the tendency for a reaction to persist after the stimulus has been withdrawn), by differences in latency of reaction when stimuli of differing strength are applied, and by the irreversibility of conduction in the reflex arc. To explain these and other discrepancies between the action of nerve trunks and the nature of the connection between stimulus and response, Sherrington postulated the functions of the synapse, the name of which he

coined. This synaptic hypothesis, based almost entirely on behavioral observations, has now been strikingly verified by advances in experimental technique. On the synaptic hypothesis, messages travel from one nerve cell to another across a small gap between the two cells. Thus there is a functional connection between the two cells as distinct from a physical fusion of their membranes. Transmission across this gap is different in kind from transmission along the processes of the nerve cell itself. (It is now known that in the vertebrate nervous system such transmission across the synapse generally occurs by means of diffusion across the gap between the nerve cells—the synaptic cleft—of small amounts of chemical substances, such as acetylcholine.) He used this hypothesis further to account for interactions between reflexes, and he also introduced the notion of the final common path, thus narrowing down the locus at which one reflex inhibits another. He provided a detailed physiological basis for the mechanism of chain reflex, a notion propounded both by Exner and by Jacques Loeb, which is of great importance to behaviorism.

It should be stressed that Sherrington, in his more than three hundred publications, made contributions to many fields other than neurophysiology, among them the physiology of blood, the psychology of contrast and flicker, and the functional localization of the motor cortex. He was also one of the early scientists to work on the neural basis of the emotions. His discussion of the James-Lange theory in the seventh chapter of *The Integrative Action of the Nervous System* is still one of the best in the literature. The excellence of his work won wide recognition, and in 1932 he was awarded the Nobel prize for medicine.

Although Sherrington probably contributed more than anyone else to the purely mechanistic understanding of reflexes, he also found it useful to consider them in terms of their purpose, or biological utility. And he was not averse to devoting thought to the more philosophical aspects of the functioning of the nervous system. In his extensive writings on the subject he argued for a Cartesian dualism of body and mind, a position in which Sir John Carew Eccles, another distinguished neurophysiologist, has followed his teacher.

In spite of their profound influence on physiology, the details of Sherrington's observations and thoughts have little affected the mainstream of behavioral science. However, the general impact of his work, whether rightly or wrongly, contributed to the vogue for analysis in terms of the reflex or of stimulus-response psychology.

[See also NERVOUS SYSTEM, especially the article on STRUCTURE AND FUNCTION OF THE BRAIN; SENSES; SENSORY AND MOTOR DEVELOPMENT; SKIN SENSES AND KINESTHESIS.]

#### WORKS BY SHERRINGTON

- (1906) 1948 *The Integrative Action of the Nervous System*. 2d ed. New Haven: Yale Univ. Press.  
 (1925) 1940 *The Assaying of Brabantius and Other Verse*. 2d ed. Oxford Univ. Press.  
 1932 CREED, R. S.; DENNY-BROWN, D.; ECCLES, J. C.; LIDDELL, E. G. T.; and SHERRINGTON, C. S. *Reflex Activity of the Spinal Cord*. Oxford: Clarendon.  
 (1940) 1951 *Man on His Nature*. 2d ed. Cambridge Univ. Press. → A paperback edition was published in 1953 by Doubleday.  
 1946 *The Endeavour of Jean Fernel*. Cambridge Univ. Press.

#### SUPPLEMENTARY BIBLIOGRAPHY

- COHEN, HENRY 1958 *Sherrington: Physiologist, Philosopher and Poet*. Springfield, Ill.: Thomas; Liverpool Univ. Press.  
 LIDDELL, EDWARD G. T. 1960 *The Discovery of Reflexes*. Oxford: Clarendon. → See especially pages 98–143, "Sherrington and His Times."

### SIDGWICK, HENRY

Henry Sidgwick (1838–1900), English economist, does not have an obvious place in the standard histories of economic or social thought. His name is not associated with any particular theory or policy, and much of his influence was exerted through his teaching and his participation in the affairs of Cambridge University.

Sidgwick was born at Skipton in Yorkshire, where his father was an Anglican clergyman and headmaster of the local grammar school. From his earliest years it was clear that Sidgwick was destined for academic success. After attending Rugby, he went to Cambridge and achieved distinction there, both in classics and mathematics. In 1859 he was elected to a fellowship at Trinity College; he remained at Cambridge for the rest of his life.

At the beginning of his academic career Sidgwick taught classics, but he was soon drawn into the field of social science and in 1867 exchanged his lectureship in classics for a lectureship in moral sciences (what would now be called social sciences). Much of his intellectual effort as a teacher at Cambridge was directed toward improving the curriculum: he was influential in raising the status of the social sciences, and he helped pave the way for the economics tripos. Thus, he may be regarded as one of the pioneers who facilitated the emergence of the "Cambridge school" of economics. He also furthered the cause of univer-

sity education for women (his wife became principal of Newnham College in 1892), and he played a part in the movement for greater religious toleration at the university.

His writings ranged over a number of fields: in addition to his early classical studies, he wrote on politics, philosophy, and economics and, more remote from his central concerns, attempted to make research on psychic phenomena (then a particular vogue in England) scientifically respectable. All who came in contact with Sidgwick seem to have agreed about his intellectual and moral stature; Marshall spoke of him as his "spiritual father and mother," and John Neville Keynes described him as the most intellectually gifted man he had ever met.

Sidgwick's written contributions to economics are to be found mainly in his *Principles of Political Economy* (1883). This work is a useful benchmark by which to judge the state of the discipline in the period between J. S. Mill's *Principles* (1848) and the first edition of Marshall's *Principles* (1890). The text is primarily a re-exposition of Mill, with traces of the emergent marginalism of Jevons and Marshall. Like Marshall, Sidgwick had a strong desire to emphasize continuity in economic thinking, rather than his own originality. As he put it, "The special aim . . . is to eliminate needless polemics by a guarded restatement of traditional doctrines, with due recognition of the advances made in economic theory by recent writers" ([1883] 1901, p. ix). The book contains some corrections and clarifications of Mill's analysis (for example, the distinction between movements along a demand schedule and shifts from one demand schedule to another, which is so blurred in Mill, is clearly made and explained by Sidgwick), and the influence of Marshall is obvious.

The most important sections in Sidgwick's *Principles* are those dealing with methodology and the analysis of government interference. Sidgwick restated the methodological position, first enunciated within the context of economics by Senior and Mill, whereby the "science" and the "art" of political economy are to be logically differentiated: "The first gives information as to what happens, without pronouncing whether it is good or bad; the second judges that what happens or would happen under certain conditions is the best thing that could happen" (*ibid.*, p. 36). However, Sidgwick (ever a moderate) refused to adopt the extreme position of Cairnes, and later of Robbins, according to which the economist qua economist has no special qualifications in the normative area of his studies.

The core of Sidgwick's major contribution is



contained in his discussion of the art of political economy (book 3). He was at pains to dissociate economic science from the political doctrines of laissez-faire and suggested various conditions under which state interference in the productive process may become desirable: one important condition is the divergence of private and social benefits. Here Sidgwick's analysis constitutes the prototype of the Marshall-Pigou approach. However, he did not present a general a priori case for government interference whenever this divergence obtains—each situation has to be considered on its own. He stated the case for interference with the distribution of income—necessitated by an unregulated economic system—with much more generality, except for reservations about the possible effects of interference on production. He argued for much greater equality of income, to be achieved by deliberate government fiscal policy, basing his argument on the "general acceptance" of two propositions of Bentham's, which may be summarized as follows: utility increases as wealth increases, but at a decreasing rate. Sidgwick did not make clear the methodological status of these propositions; we are not told whether they are judgments or testable (and tested) hypotheses.

Sidgwick's influence on social thought and policy is difficult to assess. Although he expressed grave doubts about the feasibility of a total socialist solution, his rejection of any general theoretical reasons for nonintervention and his emphasis on the need to look at every case individually place him squarely in the English Fabian tradition.

BERNARD CORRY

[For the historical context of Sidgwick's work, see the biographies of MILL and SENIOR; for discussion of the subsequent development of Sidgwick's ideas, see the biographies of MARSHALL and PIGOU.]

#### WORKS BY SIDGWICK

- (1874) 1901 *The Methods of Ethics*. 6th ed. London: Macmillan.  
 (1883) 1901 *The Principles of Political Economy*. 3d ed. New York: Macmillan.  
 (1891) 1908 *The Elements of Politics*. 4th ed. London: Macmillan.

#### SUPPLEMENTARY BIBLIOGRAPHY

- KEYNES, JOHN NEVILLE 1900 Henry Sidgwick [obituary]. *Economic Journal* 10:585-591.  
 KEYNES, JOHN NEVILLE (1908) 1963 Henry Sidgwick. Volume 3, pages 757-759 in Robert Palgrave, *Palgrave's Dictionary of Political Economy*. New York: Macmillan.  
 MARSHALL, ALFRED (1890) 1961 *Principles of Economics*. 9th ed. 2 vols. New York and London: Macmillan.

MILL, JOHN STUART (1848) 1965 *Principles of Political Economy, With Some of Their Applications to Social Philosophy*. 2 vols. Edited by J. M. Robson. *Collected Works*, Vols. 2, 3. Univ. of Toronto Press.

SIDGWICK, ARTHUR; and SIDGWICK, ELEANOR 1906 *Henry Sidgwick: A Memoir*. London and New York: Macmillan.

## SIEGFRIED, ANDRÉ

André Siegfried (1875-1959), French geographer and political scientist, gained world-wide recognition through his development of a new approach to the study of political parties known as *géographie de l'opinion politique*, or political ecology; through his studies of French political regimes; and through his more general studies of the United States and other countries.

Siegfried was born in Le Havre, the son of a wealthy Calvinist Alsatian merchant who served as mayor of Le Havre from 1878 to 1886 and as a member of the Chamber of Deputies from 1885 to 1922. His mother was the daughter of a pastor in the Reformed Church and came from the *département* of Ardèche, which Siegfried later chose as the subject of one of his studies in political ecology. His parents liked to entertain, and Siegfried had the opportunity, first in Le Havre and later in Paris, of meeting in his home the political elite of the Third Republic. As a young man, he accompanied his father, who was a passionate traveler and keen observer, on some of his journeys abroad, an experience which was important for his own later efforts to describe and compare foreign countries. His father, as mayor and as a deputy, initiated many social reforms, thereby setting an example for Siegfried's activities in this field.

Siegfried's interest in politics led him to run for the National Assembly. He was unsuccessful and decided instead to focus his political interest on the compilation of "a psychological and political geography of France." The first volume, *Tableau politique de la France de l'ouest*, appeared in 1913, but the series was interrupted by World War I. Siegfried did not return to his geographies of electoral behavior until after World War II. He then published *Géographie électorale de l'Ardèche sous la III<sup>e</sup> République* (1949a) but after that left further studies to his disciples, publishing only his basic ideas on this kind of inquiry.

Siegfried served as a professor at the *École Libre des Sciences Politiques* and from 1933 to 1946 held a chair in economic and political geography at the Collège de France. Apart from his writings his main influence came through his two courses at the

Collège de France, one on the geography of political parties in France and the other on the economic and political geography of the world. In 1942, during the German occupation, he became director of the Musée Social, an institute that his father had founded for the preparation of social legislation. He had had a large part in the establishment of the Université Populaire de Belleville, one of the earliest institutions for adult education of workingmen. He was also active as a journalist, and through his articles in *Figaro* during World War II "he became one of the most effective leaders of public opinion" (Jean Schlumberger in *André Siegfried 1875-1959* 1961). After World War II he became director of the Institut d'Études Politiques in Paris. He was elected to the Académie Française in 1944, the first geographer to be so honored.

Before and after World War I, Siegfried traveled in Great Britain, the United States, Canada, Australia, and New Zealand. "It was an opportunity," he wrote later, "to observe democracy as practiced with the Anglo-Saxon spirit and methods. Instead of the aggressive and somewhat negative individualism of the Latin civilization, I saw political societies based on social cooperation. At bottom it seemed to be a contrast between the Catholic and Protestant state of mind, religion leading to extraordinarily different consequences" (1930, p. vi). This experience reinforced his interest in the political institutions of his own country.

Siegfried's studies of foreign countries are comprehensive monographs in the tradition of Tocqueville. Each of them has a central theme: the books on New Zealand (1904) and Canada (1906) are concerned with the formation of new nations and the books on Britain (1931) and the United States (1927) with their changed positions in the world. The books are masterful syntheses of the economic, political, and cultural aspects of large complex societies, always presented with a historical perspective and a geographical foundation. He gave his major attention to class structure, to religious orientations and cleavages, and to race. Since he had more than once visited some of the countries he wrote about, he was able to observe important social changes directly. His wide range of knowledge in economics, political science, and history and his acquaintance with both the Latin and Anglo-Saxon worlds enabled him to make use of cross-cultural comparisons.

He did not resist the temptation of putting his impressions of national character, or as he put it, the "souls of nations," into print (1950). These essays, while not free of the generalizations and

exaggerations common to most writing of this sort, bring into focus the outstanding traits of each nation and are free of farfetched psychological conjectures.

Siegfried's writings on French political life fall into two categories: the historical studies of political regimes from the Third to the Fifth Republic and the ecological studies of political parties or tendencies. The latter include the early study of northwest France (1913) and the later one of Ardèche (1949a). These ecological studies represented a methodological innovation and were the beginning of a school of investigation that has gained adherents in many countries.

The observation that each of the major political tendencies in France has its particular strongholds which show a high degree of stability despite changing party labels led Siegfried to the conclusion that it was possible to distinguish regions of political opinion—political climates which correspond to regions in physical geography. The particular climate depends on the varying combinations of several major factors: (1) the social stratification of the regional population—whether it is egalitarian or hierarchic; (2) the religious orientation of the people—whether Roman Catholic or Protestant, and if Catholic, whether or not the political instructions of the priest are followed; (3) the past—that is, memories of important political events that influence political attitudes; (4) the extent to which the population is autochthonous or contains a significant proportion of "strangers" (in-migrants); (5) the geography—that is, the soil, the climate, the altitude, the topography, the settlement patterns, the population density, and the degree of accessibility or isolation of the area.

Siegfried's technique consisted essentially, although not exclusively, of comparing successive maps showing election returns by small areas with maps showing the geographic distribution of other factors and in assessing, from their fit, the relative importance of the various factors. For Ardèche, for example, the topographical map gave the best fit. In Ardèche ways of life vary with the altitude; the mountain, the hillside, and the plains communities are distinctly different. The merit of these ecological studies derives from Siegfried's intensive knowledge of the regions as well as from his sense of what is politically relevant. Although he used such formal notions as "the Right" and "the Left," he was careful to state what these mean in terms of immediate goals and general political ideas.

Although Siegfried held a chair of political and economic geography, he was not a geographer in the technical sense. The findings of geography and



geology were for him data essential to the understanding of society. Nor was he a geographical determinist, although he appreciated the opportunities and limitations which the physical environment offers to human action, particularly for the development of primary industries. The social structure, especially the class structure in the Marxian sense, depends on the use of these opportunities, and the class structure in combination with other factors produces certain political climates. But Siegfried was not a Marxist in his political orientation, or even a socialist. As a scholar, he felt obliged to avoid association with any political party. And yet, he was clearly a republican, a liberal, and a humanist and therefore also a social reformer.

While he had sympathy for the Anglo-Saxon style of political life, he regretted the "Americanizing" tendencies in the economic organization of France, the impending decline of the independent entrepreneur, and the replacement of the artisan-craftsman by the assembly-line worker. These tendencies, inescapable as they might be, he feared would endanger the dignity and spirituality of man.

RUDOLF HEBERLE

[See also GEOGRAPHY.]

#### WORKS BY SIEGFRIED

- (1904) 1914 *Democracy in New Zealand*. London: Bell. → First published in French.
- (1908) 1949 *Canada: An International Power*. 2d ed. New York: Duell. → First published as *Le Canada: Les deux races*.
- 1913 *Tableau politique de la France de l'ouest sous la Troisième République*. Paris: Colin.
- 1927 *America Comes of Age: A French Analysis*. New York: Harcourt. → First published in 1927 as *Les États-Unis d'aujourd'hui*.
- 1930 *France: A Study in Nationality*. New Haven: Yale Univ. Press. → Also published in the same year as *Tableau des partis en France*.
- 1931 *England's Crisis*. New York: Harcourt. → Also published in the same year as *La crise britannique au XX<sup>e</sup> siècle*.
- 1934 *Amérique Latine*. Paris: Colin.
- 1937 *Une géographie de l'opinion politique est-elle possible? Nouvelle revue française* 49:789-802.
- 1946 *La géographie de l'opinion politique*. *Atomes* 1, no. 4:3-6.
- 1947 *The Mediterranean*. New York: Duell.
- 1949a *La géographie électorale de l'Ardèche sous la III<sup>e</sup> République*. Paris: Colin.
- (1949b) 1950 *African Journey*. London: Cape. → First published as *Afrique du Sud: Notes de voyage*.
- (1950) 1952 *Nations Have Souls*. New York: Putnam. → First published as *L'âme des peuples*.
- 1951 *Voyage aux Indes*. Paris: Colin.
- 1952a *Géographie politique des cinq continents*. Paris: Passerell.

- 1952b *Mes souvenirs de la Troisième République; Mon père et son temps: Jules Siegfried, 1836-1922*. Paris: Presses Universitaires de France.
- (1954) 1955 *America at Mid-century*. New York: Harcourt. → First published as *Tableau des États-Unis*.
- 1956 *De la III<sup>e</sup> à la IV<sup>e</sup> République*. Paris: Grasset.
- 1958 *De la IV<sup>e</sup> à la V<sup>e</sup> République au jour le jour*. Paris: Grasset.

#### SUPPLEMENTARY BIBLIOGRAPHY

- André Siegfried 1875-1959. 1961 Paris: Imprimerie Foulon. → A collection of eulogies. See especially the article by Jean Schlumberger.

## SIGNIFICANCE, TESTS OF

*The topic of significance testing is treated in detail in HYPOTHESIS TESTING, a term that is usually regarded as synonymous with "tests of significance," although some statisticians feel that there are important distinctions. The present article describes the basic ideas of significance testing, outlines the most important elementary tests, and reviews problems related to the philosophy and application of significance testing.*

Significance tests are statistical procedures for examining a hypothesis in the light of observations. Many significance tests are simple and widely used, and have a long history; the idea of significance testing goes back at least to the eighteenth century. There has been much confusion about significance testing, and consequently it has been much misused. Despite their apparent simplicity (and sometimes because of it), significance tests have generated many controversies about meaning and utility.

A significance test starts with observations and with a hypothesis about the chance mechanism generating the observations. From the observations a test statistic is formed. Large values of the test statistic (or small values, or both, depending on circumstances) lead to strong skepticism about the hypothesis, whereas other values of the test statistic are held to be in conformance with the hypothesis. Choice of the test statistic usually depends upon the alternatives to the hypothesis under test.

### Basic ideas

As an example of significance testing, suppose that the hypothesis under test, often called the *null hypothesis*, or  $H_0$ , is that sleep deprivation has no effect on a certain kind of human skill, say arithmetic ability as measured by a particular type of test. Some population must be specified; suppose

that it consists of all students at a particular college who are willing to participate in a sleep-deprivation experiment. One hundred students are chosen at random, tested for arithmetic skill, deprived of a night's sleep, and then tested again. For each student the initial test result minus the second test result is regarded as the basic observation, datum, or score. Suppose, for the present, that if the null hypothesis is false, the effect of sleep deprivation is either to increase average score or to decrease it; either direction might hold, so the *alternative hypotheses* are two-sided.

The analyst would, of course, be concerned about possible practice and motivational effects; a helpful device might be the use of a control group that would also be tested twice, but with a normal night's sleep. The effect of practice might be lessened by the presentation of a sequence of tests before the experiment proper begins. Another matter of concern is whether the basic score used, the *difference* between test results, is more appropriate than the ratio of test results or some other combination of them. For present purposes such issues are not discussed.

From the 100 scores a test statistic is then formed. The choice might well be the Student *t*-statistic, which is the observed average score divided by an estimate of its own variability. Details of the *t*-statistic are given later.

The nub of the significance-testing concept is that if the null hypothesis really is true, then it is unlikely that the Student *t*-statistic would be greater than, say, 2.6 in unsigned numerical value. The probability of that event, under the usual—but possibly dangerous—assumptions that the scores behave like statistically independent random variables with the same normal distribution, is about .01, or 1 out of 100. If, on the other hand, the null hypothesis is false, the probability of the event is greater than .01, much greater if sleep deprivation has a large effect.

Hence, if a Student *t*-statistic greater than 2.6 or less than -2.6 is observed, *either* the null hypothesis is false, *or* the null hypothesis is true and an event of small probability has occurred, *or* something is wrong with the underlying assumptions. If those assumptions seem satisfactory, and if the Student *t*-statistic is, say, 2.7, most people would prefer to act as if the null hypothesis were false. The cut-off number, 2.6, and the associated probability, .01, are cited for the sake of specific illustration. If 2.6 is replaced by a larger number, the associated probability is smaller; if it is replaced by a smaller number, the probability is larger.

In formal significance testing one decides beforehand on the cut-off number. If 2.6 is chosen, then there is a .01 probability of erroneously doubting the null hypothesis when it is in fact true.

In the example, 2.6 is called the *critical value* of the (unsigned) Student *t*-statistic, and .01 is called the *level of significance* or (usually synonymously) the *size* of the test. The level of significance, frequently denoted by  $\alpha$ , is also called the probability of Type I error, or of error of the first kind; this is the error of falsely doubting the null hypothesis when it is in fact true. In the present context, if the unsigned Student *t*-statistic turns out to be greater than the critical value, it is often said that the statistic (or the test, or the sample) is *statistically significant*; this terminology can lead to confusion, as will be described below.

From this approach, then, a significance test is a procedure, agreed upon beforehand, that will result in one of two conclusions (or actions, or viewpoints): if the test statistic—here the unsigned Student *t*-statistic—is greater than the critical value, one tends to act as if the null hypothesis is false; otherwise one tends to act in some different way, not necessarily as if the null hypothesis is true but at least as if its falsity is not proved. These two kinds of conclusions or actions are conventionally called, respectively, *rejection* and *acceptance* of the null hypothesis. Such nomenclature, with its emotional overtones and its connotations in everyday speech, has led to confusion and needless controversy. The terms "rejection" and "acceptance" should be thought of simply as labels for two kinds of actions or conclusions, the first more appropriate if the null hypothesis is false and the second more appropriate if it is true or at least not known to be false. In the example, rejection and acceptance might correspond to different attitudes toward a theory of sleep or to different recommendations for desirable amounts of sleep. (In some contexts, such as acceptance sampling in industry, the terms "acceptance" and "rejection" may be used literally.)

There is further room for confusion in the varying ways that significance testing is described. In the example it would be accurate to say that the hypothesis of no sleep-deprivation effect (on average score) is being tested. But it might also be said, more loosely, that one is testing the effect of sleep deprivation, and this might be misinterpreted to mean that the null hypothesis asserts a positive mean score under sleep deprivation. One reason for this ambiguity is that the null hypothesis is often set up in the hope that it will be rejected. In the comparison of a new medical treatment with an old one, equality of effectiveness might be un-



interesting, so only rejection of the null hypothesis of equality would excite the experimenter, especially if he had invented the new treatment. On the other hand, accepting the null hypothesis may itself be very important—for example, in an experiment to test the theory of relativity. Usually, however, there is a basic asymmetry between the null hypothesis and the other hypotheses.

The two-decision procedure discussed above is a highly simplified framework for many situations in which significance tests are used. Although there are cases in which the above viewpoint makes direct sense (for example, industrial acceptance sampling, crucial experiments for a scientific theory), most uses of significance tests are different. In these more common uses the analyst computes that level of significance at which the observed test statistic would just lead to rejection of the null hypothesis. In the example, if the observed Student *t*-statistic were 1.8, the resulting *sample level of significance*, or *observed significance level*, would be .075, again under the conventional assumptions listed earlier. From this second viewpoint the observed (or sample) significance level is a measure of rarity or surprise; the smaller the level, the more surprising the result if the null hypothesis is true. Naturally, if the result is very surprising under the null hypothesis, one is likely to conclude that the null hypothesis is false, but it is important to know quantitatively the magnitude of the surprise. The sample level of significance is often called the *P*-value, and the notation  $P = .075$  (or  $p = .075$ ) may be used. (For other views on measuring surprise, see Weaver 1948; Good 1956.)

Usually a very rough knowledge of the degree of surprise suffices, and various simplifying conventions have come into use. For example, if the observed significance level lies between .05 and .01, some authors say that the sample is *statistically significant*, while if it is less than .01, they call it *highly statistically significant*. The adverb "statistically" is often omitted, and this is unfortunate, since statistical significance of a sample bears no necessary relationship to possible subject-matter significance of whatever true departure from the null hypothesis might obtain (see Boring 1919).

From one point of view, significance testing is a device for curbing overenthusiasm about apparent effects in the data. For example, if the sleep-deprivation experiment showed a positive effect, an experimenter enthusiastic about a theory of sleep psychology might cry, "Eureka! Sleep deprivation really does have the predicted consequence." After he himself has had a good night's sleep, however, he might compute the appropriate test statistic and see, perhaps, that the observed effect—

or a greater one—could occur with probability .2 under the null hypothesis of no real underlying effect. This would probably dampen his initial enthusiasm and might prevent one of the many premature announcements of discoveries that clutter the literature of some sciences. (The sleep example is solely for illustration; I have no reason to think that experimenters on sleep are anything but cautious in their statistical analyses.)

On the other hand, it is easy to be overconservative and throw out an interesting baby with the nonsignificant bath water. Lack of statistical significance at a conventional level does not mean that no real effect is present; it means only that no real effect is clearly seen from the data. That is why it is of the highest importance to look at power and to compute confidence intervals, procedures that will be discussed later.

The null hypothesis, despite its name, need not be that the average effect of sleep deprivation is zero. One might test the null hypothesis that the average effect is 10 score units. Or one might test such a null hypothesis as: the average effect of sleep deprivation is not positive; or, again: the average effect lies between -10 and 10 score units. The last two null hypotheses, unlike the previous ones, do not fix the hypothesized average effect at a single value.

### Standard tests

There are a number of popular standard significance tests in the statistical armamentarium. The tests described here include those on means, variances, correlations, proportions, and goodness of fit.

**Mean.** The null hypothesis in a test on means, as in the original example, is that the expected value (mean) of a distribution is some given number,  $\mu_0$ . This hypothesis is to be tested on the basis of a random sample of  $n$  observations on the distribution; that is, one considers  $n$  independent random variables, each with that same distribution whose mean is of interest. If the variance,  $\sigma^2$ , of the distribution is known, and if  $\bar{X}$  denotes the sample mean (arithmetic average), a widely useful test statistic is

$$(1) \quad \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}},$$

which has mean zero and standard deviation one under the null hypothesis. Often the absolute (unsigned numerical) value of (1) is the test statistic. If the underlying distribution of interest is normal, then (1) is itself normally distributed, with variance one and mean zero (that is, it has a unit normal distribution) under the null hypothesis.

Even if the underlying distribution is not normal, (1) has approximately the unit normal distribution under the null hypothesis, provided that  $n$  is not too small.

In practice  $\sigma^2$  is usually unknown, and an estimator of it,  $s^2$ , is used instead. The most usual estimator is based on the sum of squared deviations of the observations from  $\bar{X}$ ,

$$(2) \quad s^2 = \frac{1}{n-1} [(X_1 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2],$$

where the  $X_i$  are the observations. (The usual name for  $s^2$  is "sample variance," and that for  $s$  is "sample standard deviation.") Then (1) becomes the Student (or  $t$ -) statistic

$$(3) \quad \frac{\bar{X} - \mu_0}{s/\sqrt{n}}.$$

If the  $X_i$  are normal, then (3) has as its distribution under the null hypothesis the Student (or  $t$ -) distribution with  $n-1$  degrees of freedom ( $df$ ). If  $n$  is large (and often it need not be very large), then in quite general circumstances the distribution of (3) under the null hypothesis is approximately the unit normal distribution. [See DISTRIBUTIONS, STATISTICAL, articles on SPECIAL CONTINUOUS DISTRIBUTIONS and APPROXIMATIONS TO DISTRIBUTIONS.]

*Two samples.* Suppose that the observations form two independent random samples of sizes  $n_1$  and  $n_2$ :  $X_{11}, \dots, X_{1n_1}$  and  $X_{21}, \dots, X_{2n_2}$ . Here the first subscript denotes the sample (1 or 2), and the second subscript denotes the observation within the sample (1 through  $n_1$  or  $n_2$ ). For example, in the sleep-deprivation experiment the first sample might be of male students and the second of female students. The null hypothesis might be that the average effect of sleep deprivation on arithmetic skill is the same for male and female students. Note that this null hypothesis says nothing about the magnitude of the average effect; it says only that the average effect is the same for the two groups.

More generally, suppose the null hypothesis is that the two underlying means are equal:  $\mu_1 = \mu_2$ , where  $\mu_1$  is the expected value of any of the  $X_{1i}$  and  $\mu_2$  that of any of the  $X_{2i}$ . If the variances are known, say  $\sigma_1^2$  and  $\sigma_2^2$ , the usual test statistic is

$$(4) \quad \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}},$$

where  $\bar{X}_1$  and  $\bar{X}_2$  are the two sample means. If the  $X$ 's are normally distributed, (4) has the unit normal as null distribution—that is, as its distribution when the null hypothesis holds. Even if the  $X$ 's are

not normally distributed, the unit normal distribution is a good approximation to the null distribution of (4) for  $n_1, n_2$  large. (If, instead of the null hypothesis  $\mu_1 = \mu_2$ , one wishes to test  $\mu_1 = \mu_2 + \delta$ , where  $\delta$  is some given number, one need only replace  $\bar{X}_1 - \bar{X}_2$  in the numerator of (4) with  $\bar{X}_1 - \bar{X}_2 - \delta$ .)

If the variances are not known, and if  $s_1^2$  and  $s_2^2$  are estimators of them, a common test statistic is

$$(5) \quad \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}},$$

which, for  $n_1, n_2$  large, has the unit normal distribution as an approximation to the null distribution. The null distribution is better approximated by a  $t$ -distribution [details can be found in ERRORS, article on EFFECTS OF ERRORS IN STATISTICAL ASSUMPTIONS].

If the variances are unknown but may reasonably be assumed to be equal ( $\sigma_1^2 = \sigma_2^2$ ), then a common test statistic is the two-sample Student statistic,

$$(6) \quad \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}},$$

where  $s_1^2, s_2^2$  are defined in terms of (2). If the observations are normally distributed, (6) has as null distribution the  $t$ -distribution with  $n_1 + n_2 - 2$   $df$ . This holds approximately, in general, if the  $X$ 's are mildly nonnormally distributed and if  $\sigma_1^2$  and  $\sigma_2^2$  differ somewhat but  $n_1$  and  $n_2$  are not very different.

*Paired samples.* The two-sample procedure described above should be distinguished from the paired two-sample procedure. The latter might be used if the subjects are fraternal twins of opposite sex, with both members of a sampled pair of twins observed. Then the male and female samples would no longer be independent, since presumably the members of a pair of twins are more nearly alike in most respects than are two random subjects. In this case one would still use  $\bar{X}_1 - \bar{X}_2$  in the numerator of the test statistic, but the denominator would be different from those given above. Although it is not always possible, pairing (more generally, blocking) is often used to increase the power of a test. The underlying idea is to make comparisons within a relatively homogeneous set of experimental units. [See EXPERIMENTAL DESIGN.]

[These significance tests, and generalizations of them, are further discussed in HYPOTHESIS TESTING; LINEAR HYPOTHESES, article on ANALYSIS OF VARIANCE; and ERRORS, article on EFFECTS OF ERRORS IN STATISTICAL ASSUMPTIONS.]



**Variances.** The null hypothesis in a test on variances may be that  $\sigma^2$  has the value  $\sigma_0^2$ ; then the usual test statistic is  $s^2$  as defined in (2), or, more conveniently,

$$(7) \quad (n-1)s^2/\sigma_0^2.$$

Under the null hypothesis, (7) has the chi-square distribution with  $n-1$  degrees of freedom, provided that the observations are normal, independent, and identically distributed. Unlike the null distributions for the tests on means, the null distribution of (7) is highly sensitive to deviations from normality.

In the two-sample case a common null hypothesis is  $\sigma_1^2 = \sigma_2^2$ . The usual test statistic is  $s_1^2/s_2^2$ , which, under normality, has as null distribution the  $F$ -distribution with  $n_1 - 1$  and  $n_2 - 1$  *df*.

[These tests are discussed further in LINEAR HYPOTHESES, article on ANALYSIS OF VARIANCE; VARIANCES, STATISTICAL STUDY OF.]

**Correlation.** The most common procedure in simple correlation analysis is to test the null hypothesis that the population correlation coefficient is zero, using as test statistic the sample correlation coefficient. Under the assumptions of bivariate normality and a random sample, the null distribution is closely related to a  $t$ -distribution. One can also test the null hypothesis that the population correlation coefficient has some nonzero value, say .55. Special tables or approximate methods are required here [see MULTIVARIATE ANALYSIS, articles on CORRELATION].

**Proportions. Single sample.** The simplest case in the analysis of proportions is that of a sample proportion and the null hypothesis that the corresponding population probability is  $p_0$ , some given number between 0 and 1. The sample proportion and population probability correspond to some dichotomous property: alive-dead, heads-tails, success-failure. It is convenient to summarize the sample in the form of a simple table

$N$	$n - N$	$n$
-----	---------	-----

where  $n$  is the sample size,  $N$  the number of sample observations having a stated property,  $n - N$  the number not having this property, and  $N/n$  the sample proportion. The usual sampling assumptions, which need examination in each application, are that the observations are statistically independent and that each has the same probability of having the stated property. As a result of these assumptions,  $N$  has a binomial distribution.

The usual test statistic is

$$(8) \quad \frac{N - np_0}{\sqrt{np_0(1-p_0)}},$$

which, for  $n$  not too small, has approximately the unit normal distribution under the null hypothesis. The square of this test statistic, again for  $n$  not too small and under the null hypothesis, has approximately a chi-square distribution with one *df*.

Notice that (8) is really a special case of (1) if  $N$  is regarded as the sum of  $n$  independent observations taking values 1 and 0 with probabilities  $p_0$  and  $1 - p_0$ .

**Two samples.** The data may consist of two sample proportions, and the null hypothesis may be that the corresponding population proportions are equal. It is convenient to express such data in terms of a  $2 \times 2$  table like Table 1, where the two samples (of sizes  $n_{1+}$ ,  $n_{2+}$ ) correspond to the upper two lines and the bottom and right rims are marginal totals—for example,  $N_{11} + N_{12} = n_{1+}$ . Capital letters denote random variables and lower-case letters nonrandom quantities. The usual assumptions, which must be examined in each application, are that  $N_{11}$  and  $N_{21}$  are independently and binomially distributed. A common test statistic is

$$(9) \quad \frac{(N_{11}/n_{1+}) - (N_{21}/n_{2+})}{\sqrt{\left(\frac{1}{n_{1+}} + \frac{1}{n_{2+}}\right) \frac{N_{+1}N_{+2}}{n^2}}},$$

which, for  $n$ 's not too small, has approximately the unit normal distribution under the null hypothesis. (Note the great similarity of test statistics (6) and (9), which are related in much the same way as (1) and (8).)

If the unsigned value of (9) is the test statistic, one may equivalently consider its square, which is expressible in the simpler form

$$(10) \quad n \frac{(N_{11}N_{22} - N_{12}N_{21})^2}{n_{1+}n_{2+}N_{+1}N_{+2}}.$$

Under the null hypothesis, and for  $n$ 's not too small, (10) has approximately a chi-square distribution with one *df*. Large values of (10) are statistically significant.

**Chi-square statistics.** The above test statistics, in squared form, are special cases of the chi-square test statistic; their null distributions are approxi-

Table 1 — Summary of two samples for analysis of proportions

	Number having property	Number not having property	Totals
Sample 1	$N_{11}$	$N_{12}$	$n_{1+}$
Sample 2	$N_{21}$	$N_{22}$	$n_{2+}$
Totals	$N_{+1}$	$N_{+2}$	$n_{++} = n$

mately chi-square ones. Such test statistics can generally be thought of in terms of data tabulations of the kind shown above; the statistic is the sum of terms given by the mnemonic expression

$$\frac{(\text{observed} - \text{"expected"})^2}{\text{"expected"}},$$

summed over the cells in the body of the table. For example, in the first, very simple table for a sample proportion, there are two cells in the body of the table, with  $N$  and  $n - N$  observed frequencies in them. The corresponding frequencies expected under the null hypothesis are  $np_0$  and  $n(1 - p_0)$ . Applying the mnemonic expression gives as the chi-square statistic

$$\frac{(N - np_0)^2}{np_0} + \frac{[n - N - n(1 - p_0)]^2}{n(1 - p_0)},$$

which can readily be shown to equal the square of (8). The one  $df$  for the approximate chi-square distribution is sometimes described thus: There are *two* observed frequencies, but they are restricted in that their sum must be  $n$ . Hence, there is *one*  $df$ .

Similarly, (10) can be obtained from the table for two proportions by means of the mnemonic expression. In this case the null hypothesis does not completely specify the expected frequencies, so they must be estimated from the data. That is why quotation marks appear around "expected" in the mnemonic. [Further details, and related procedures, including the useful continuity correction, are discussed in COUNTED DATA.]

**Association for counted data.** Suppose that the individuals in a sample of people are cross-classified by hair color and eye color, with perhaps five categories for each attribute. The null hypothesis is that the two attributes are statistically independent. A standard procedure to test for association in this *contingency table* is another chi-square test, which is based essentially on (10) above. Although the test statistics are the same and have the same approximate null distribution, the power functions are different. [This test and related ones are discussed in COUNTED DATA.]

**Nonparametric tests.** The tests for means, variances, and correlation, discussed above, are closely tied to normality assumptions, or at least to approximate normality. Analogous tests of significance, without the normality restriction, have been devised [these are discussed in NONPARAMETRIC STATISTICS; HYPOTHESIS TESTING].

**General approximation.** In many cases a test statistic,  $T$ , is at hand, but its null distribution is difficult to express in usable form. It is often useful to approximate this null distribution in terms of  $\mu_0$ , the expected value of  $T$  under the null hypothesis,

and  $S$ , an estimator of the standard deviation of  $T$  under the null hypothesis. The approximation takes  $(T - \mu_0)/S$  as roughly unit normal under the null hypothesis. This approximation is, intuitively and historically, very important. The first tests of proportions discussed above are special cases of the approximation.

**Goodness of fit.** A number of significance tests are directed to the problem of goodness of fit. In these cases the null hypothesis is that the parent distribution is some specific one, or a member of some specific family. Another problem often classed as one of goodness of fit is that of testing, on the basis of two samples, the hypothesis that two parent populations are the same. [See GOODNESS OF FIT.]

### Other aspects of significance testing

**Alternative hypotheses and power.** A significance test of a null hypothesis is usually held to make sense only when one has at least a rough idea of the hypotheses that hold if the null hypothesis does not; these are the *alternative hypotheses* (or *alternate hypotheses*). In testing that a population mean,  $\mu$ , is zero, as in the example at the start of the article, the test is different if one is interested only in positive alternatives ( $\mu > 0$ ), if one is interested only in negative alternatives ( $\mu < 0$ ), or if interest extends to both positive and negative alternatives.

When using such a test statistic as the Student  $t$ -statistic, it is important to keep the alternatives in mind. In the sleep-deprivation example the original discussion is appropriate when both positive and negative alternatives are relevant. The null hypothesis is rejected if the test statistic has either a surprisingly large positive value or a surprisingly large negative value. On the other hand, it might be known that sleep deprivation, if it affects arithmetic skill at all, can only make it poorer. This means that the expected average score (where the score is the initial test result minus the result after sleep deprivation) cannot possibly be negative but must be zero or positive. One would then use the so-called right-tail test, rejecting the null hypothesis only if the Student  $t$ -statistic is large. For example, deciding to reject  $H_0$  when the statistic is greater than 2.6 leads to a level of significance of .005. Similarly, right-tail  $P$ -values would be computed. If the Student statistic observed is 1.8, the right-tail  $P$ -value is .037; that is, the probability, under the null hypothesis, of observing a Student statistic 1.8 or larger is .037. There are also left-tail tests and left-tail  $P$ -values.

For a test considered as a two-action procedure, the *power* is the probability of (properly) rejecting the null hypothesis when it is false. Power, of



course, is a function of the particular alternative hypothesis in question, and *power functions* have been much studied. The probability of error of Type II is one minus power; a Type II error is acceptance of the null hypothesis when it is false. By using formulas, approximations, tables, or graphs for power, one can determine sample sizes so as to control both size and power. [See EXPERIMENTAL DESIGN, *article on THE DESIGN OF EXPERIMENTS*.]

A common error in using significance tests is to neglect power considerations and to conclude from a sample leading to "acceptance" of the null hypothesis that the null hypothesis holds. If the power of the test is low for relevant alternative hypotheses, then a sample leading to acceptance of  $H_0$  is also very likely if those alternatives hold, and the conclusion is therefore unwarranted. Conversely, if the sample is very large, power may be high for alternatives "close to" the null hypothesis; it is important that this be recognized and possibly reacted to by decreasing the level of significance. These points are discussed again later.

It is most important to consider the power function of a significance test, even if only crudely and approximately. As Jerzy Neyman wrote, perhaps with a bit of exaggeration, "... if experimenters realized how little is the chance of their experiments discovering what they are intended to discover, then a very substantial proportion of the experiments now in progress would have been abandoned in favour of an increase in size of the remaining experiments, judged more important" (1958, p. 15). [References to Neyman's *fundamental and pathbreaking contributions to the theory of testing* are given in HYPOTHESIS TESTING.]

There is, however, another point of view for which significance tests of a null hypothesis may be relevant without consideration of alternative hypotheses and power. An illuminating discussion of this viewpoint is given by Anscombe (1963).

**Combining significance tests.** Sometimes it is desirable to combine two or more significance tests into a single one without reanalyzing the detailed data. For example, one may have from published materials only the sample significance levels of two tests on the same hypothesis or closely related ones. (Discussions of how to do this are given in Mosteller & Bush [1954] 1959, pp. 328-331; and Moses 1956. See also Good 1958.) Of course, a combined analysis of all the data is usually desirable, when that is possible.

**Preliminary tests of significance.** A desired test procedure is often based on an assumption that may be questionable—for example, equality of the

two variances in the two-sample  $t$ -test. It has frequently been suggested that a preliminary test of the assumption (as null hypothesis) be made and that the test of primary interest then be carried out only if the assumption is accepted by the preliminary test. If the assumption is rejected, the test of primary interest requires modification.

Such a two-step procedure is difficult to analyze and must be carried out with caution. One relevant issue is the effect on the primary test of a given error in the assumption. Another is that the preliminary test may be much more sensitive to errors in other underlying assumptions than is the main significance test. [See ERRORS, *article on EFFECTS OF ERRORS IN STATISTICAL ASSUMPTIONS*. A discussion of preliminary significance tests, with references to prior literature, is given in Bancroft 1964. Related material is given in Kitagawa 1963.]

**Relation to confidence sets and estimation.** It often makes sense, although the procedure is not usually described in these terms, to compute appropriate sample significance levels not only for the null hypothesis but also for alternative hypotheses as if—for the moment—each were a null hypothesis. In this way one obtains a measure of how surprising the sample is for both the null and the alternative hypotheses. If parametric values corresponding to those hypotheses not surprising (at a specified level) for the sample are considered together, a *confidence region* is obtained. [See ESTIMATION, *article on CONFIDENCE INTERVALS AND REGIONS*.]

In any case, a significance test is typically only one step in a statistical analysis. A test asks, in effect, Is anything other than random variation appearing beyond what is specified by the null hypothesis? Whatever the answer to that question is—but especially if the answer is Yes—it is almost always important to estimate the magnitudes of underlying effects. [See ESTIMATION.]

**Relation to discriminant analysis.** Significance testing, historically and in most presentations, is asymmetrical: control of significance level is more important than control of power. The null hypothesis has a privileged position. This is sometimes reasonable—for example, when the alternative hypotheses are diffuse while the null hypothesis is sharp. In other cases there is no particular reason to call one hypothesis "null" and the other "alternative" and hence no reason for asymmetry in the treatment of the two kinds of error. This symmetric treatment then is much the same as certain parts of the field called discriminant analysis. [See MULTIVARIATE ANALYSIS, *article on CLASSIFICATION AND DISCRIMINATION*.]

### Dangers, problems, and criticisms

Some dangers and problems of significance testing have already been touched on: failure to consider power, rigid misinterpretation of "accept" and "reject," serious invalidity of assumptions. Further dangers and problems are now discussed, along with related criticisms of significance testing.

**Nonsignificance is often nonpublic.** Negative results are not so likely to reach publication as are positive ones. In most significance-testing situations a negative result is a result that is not statistically significant, and hence one sees in published papers and books many more statistically significant results than might be expected. Many—perhaps most—statistically nonsignificant results never see publication.

The effect of this is to change the interpretation of published significance tests in a way that is hard to analyze quantitatively. Suppose, to take a simple case, that some null hypothesis is investigated independently by a number of experimenters, all testing at the .05 level of significance. Suppose, further, that the null hypothesis is true. Then any one experimenter will have only a 5/100 chance of (misleadingly) finding statistical significance, but the chance that *at least one* experimenter will find statistical significance is appreciably higher. If, for example, there are six experimenters, a rejection of the null hypothesis by at least one of them will take place with probability .265, that is, more than one time out of four. If papers about experiments are much more likely to be published when a significance test shows a level of .05 (or less) than otherwise, then the nonpublication of nonsignificant results can lead to apparent contradictions and substantive controversy. If the null hypothesis is false, a similar analysis shows that the "power" of published significance tests may be appreciably higher than their nominal power. (Discussions of this problem are given in Sterling 1959; and Tullock 1959.)

**Complete populations.** Another difficulty arises in the use of significance tests (or any other procedures of probabilistic inference) when the data consist of a complete census for the relevant population. For example, suppose that per capita income and per capita dollars spent on new automobiles are examined for the 50 states of the United States in 1964. The formal correlation coefficient may readily be computed and may have utility as a descriptive summary, but it would be highly questionable to use sampling theory of the kind discussed in this article to test the null hypothesis that the population correlation coefficient

is zero, or is some other value. The difficulty is much more fundamental than that of nonnormality; it is hard to see how the 50 pairs of numbers can reasonably be regarded as a sample of any kind. Some statisticians believe that permutation tests may often be used meaningfully in such a context, but there is no consensus. [*For a definition of permutation tests, see NONPARAMETRIC STATISTICS. Further discussion of this problem and additional references are given in Hirschi & Selvin 1967, chapter 13. An early article is Woofter 1933.*]

**Target versus sampled populations.** Significance tests also share with all other kinds of inference from samples the difficulty that the population sampled from is usually more limited than the broader population for which an inference is desired. In the sleep-deprivation example the sampled population consists of students at a particular college who are willing to be experimental subjects. Presumably one wants to make inferences about a wider population: all students at the college, willing or not; all people of college age; perhaps all people. [*See STATISTICS; ERRORS, article on NON-SAMPLING ERRORS.*]

**Neglect of power by word play.** A fallacious argument is that power and error of the second kind (accepting the null hypothesis when it is false) need not be of concern, since the null hypothesis is never really accepted but is just not rejected. This is arrant playing with words, since a significance test is fatuous unless there is a question with at least two possible answers in the background. Hence, both kinds of probabilities of wrong answers are important to consider. Recall that "accept" and "reject" are token words, each corresponding to a conclusion that is relatively more desirable when one or another true state of affairs obtains.

To see in another way why more than Type I error alone must be kept in mind, notice that one can, without any experiment or expense, achieve zero level of significance (no error of the first kind) by never rejecting the null hypothesis. Or one can achieve any desired significance level by using a random device (like dice or a roulette wheel) to decide the issue, without any substantive experiment. Of course, such a procedure is absurd because, in the terminology used here, its power equals its level of significance, whatever alternative hypothesis may hold.

**Difficulties with significance level.** If significance tests are regarded as decision procedures, one criticism of them is based on the arbitrariness of level of significance and the lack of guidance in choosing that level. The usual advice is to examine



level of significance and power for various possible sample sizes and experimental designs and then to choose the least expensive design with satisfactory characteristics. But how is one to know what is satisfactory? Some say that if satisfaction cannot be described, then the experimenter has not thought deeply enough about his materials. Others oppose such a viewpoint as overmechanical and inappropriate in scientific inference; they might add that the arbitrariness of size, or something analogous to it, is intrinsic in any inferential process. One cannot make an omelet without deciding how many eggs to break.

*Unconventional significance levels.* Probably the most common significance levels are .05 and .01, and tables of critical values are generally for these levels. But special circumstances may dictate tighter or looser levels. In evaluating the safety of a drug to be used on human beings, one might impose a significance level of .001. In exploratory work, it might be quite reasonable to use levels of .10 or .15 in order to increase power. What is of central importance is to know what one is doing and, in particular, to know the properties of the test that is used.

*Nonconstant significance levels.* For many test situations it is impossible to obtain a sensible test with the same level of significance for all distributions described by a so-called composite null hypothesis. For example, in testing the null hypothesis that a population mean is not positive, the level of significance depends generally on the value of the nonpositive mean; the more it departs from zero, the smaller the probability of Type I error. In such cases the usual approach is to think in terms of the *maximum* probability of Type I error over all distributions described by the null hypothesis. In the nonpositive mean case, the maximum is typically attained when the mean is zero.

*Necessarily false null hypotheses.* Another criticism of standard significance tests is that in most applications it is known beforehand that the null hypothesis cannot be *exactly* true. For example, it seems most implausible that sleep deprivation should have literally no effect at all on arithmetic ability. Hence, why bother to test a null hypothesis known at the start to be false?

One answer to this criticism can be outlined as follows: A test of the hypothesis that a population mean has a specified value,  $\mu_0$ , is a simplification. What one really wants to test is whether the mean is *near*  $\mu_0$ , as near as makes no substantive difference. For example, if a new psychological therapy raises the cure rate from 51 per cent to 51.1 per cent, then even if one could discover such a small

difference, it might be substantively uninteresting and unimportant. For "reasonable" sample sizes and "reasonable" significance levels, most standard tests have power quite close to the level of significance for alternative hypotheses close to the null hypothesis. When this is so and when, in addition, power is at least moderately large for alternatives interestingly different from the null hypothesis, one is in a satisfactory position, and the criticism of this section is not applicable. The word "reasonable" is in quotation marks above because what is in fact reasonable depends strongly on context. To examine reasonableness it is necessary to inspect, at least roughly, the entire power function. Many misuses of significance testing spring from complete disregard of power.

A few authors, notably Hodges and Lehmann (1954), have formalized the argument outlined above by investigating tests of null hypotheses such as the following: The population mean is in the interval  $(\mu_0 - \Delta, \mu_0 + \Delta)$ , where  $\mu_0$  is given and  $\Delta$  is a given positive number.

There are, to be sure, null hypotheses that are not regarded beforehand as surely false. One well-known example is the null hypothesis that there is no extrasensory effect in a parapsychological experiment. Other examples are found in crucial tests of well-formulated physical theories. Even in these instances the presence of small measurement biases may make interpretation difficult. A minuscule measurement bias, perhaps stemming from the scoring method, in an extensive parapsychological experiment may give a statistically significant result although no real parapsychological effect is present. In such a case the statistical significance may reflect only the measurement bias, and thus much controversy about parapsychology centers about the nature and magnitude of possible measurement biases, including chicanery, unconscious cues, and biases of scoring.

A sharp attack on significance testing, along the lines of this section and others, is given by L. J. Savage (1954, chapter 16).

*Several tests on the same data.* Frequently, two or more hypothesis tests are carried out on the same data. Although each test may be at, say, the .05 level of significance, one may ask about the *joint* behavior of the tests. For example, it may be important to know the probability of Type I error for both of two tests together. When the test statistics are statistically independent, there is no problem; the probability, for example, that at least one of two tests at level .05 will give rise to Type I error is  $1 - (.95)^2 = .0975$ . But in general the tests are statistically dependent, and analogous compu-

tations may be difficult. Yet it is important to know, for example, when two tests are positively associated, in the sense that given a Type I error by one, the other is highly likely to have a Type I error.

When a moderate to large number of tests are carried out on the same data, there is generally a high probability that a few will show statistical significance even when all the null hypotheses are true. The reason is that although the tests are dependent, they are not completely so, and by making many tests one increases the probability that at least some show statistical significance. [*Some ways of mitigating these problems for special cases are described in LINEAR HYPOTHESES, article on MULTIPLE COMPARISONS.*]

**One-sided versus two-sided tests.** There has been much discussion in the social science literature (especially the psychological literature) of when one-sided or two-sided tests should be used. If a one-sided test is performed, with the choice of sidedness made tendentiously *after* inspection of the data, then the nominal significance level is grossly distorted; in simple cases the actual significance level is twice the nominal one.

Suppose that in the sleep-deprivation example the data show an average score that is positive. "Aha," says the investigator, "I will test the null hypothesis with a right-tail test, against positive means as alternatives." Clearly, if he pursues that policy for observed positive average scores and the opposite policy for observed negative average scores, the investigator is unwittingly doing a two-tail test with double the stated significance level. This is an insidious problem because it is usually easy to rationalize a choice of sidedness *post hoc*, so that the investigator may be fooling both himself and his audience.

The same problem occurs in the choice of test statistic in general. If six drugs are compared in their effects on arithmetic skill and only the observations on those two drugs with least and greatest observed effects are chosen for a test statistic, with no account taken of the choice in computing critical values, an apparently statistically significant result may well just reflect random error, in the sense that the true significance level is much higher than the nominal one. [*Some methods of dealing with this are described in LINEAR HYPOTHESES, article on MULTIPLE COMPARISONS.*]

**Three-decision procedures.** Many writers have worried about what to do after a two-tail test shows statistical significance. One might conclude, for example, that sleep deprivation has an effect on average score, but one wishes to go further and say that it has a positive or a negative effect, de-

pending on the sign of the sample average. Yet significance testing, regarded stringently as a two-decision procedure, makes no provision for such conclusions about sign of effect when a two-sided test rejects the null hypothesis.

What is really wanted in such a case may well be a *three-decision* procedure: one either accepts  $H_0$ , rejects  $H_0$  in favor of alternatives on one side, or rejects  $H_0$  in favor of alternatives on the other side. There is no reason why such a procedure should not be used, and it is, in effect, often used when the user says that he is carrying out a significance test. A major new consideration with this kind of three-decision procedure is that one now has *six*, rather than *two*, kinds of error whose probabilities are relevant; for each possible hypothesis there are two erroneous decisions. In some simple cases symmetry reduces the number of different probabilities from six to three; further, some of the probabilities may be negligibly small, although these are probabilities of particularly serious errors (for a discussion, see Kaiser 1960). A three-decision procedure may often be usefully regarded as a composition of two one-sided tests. [*See HYPOTHESIS TESTING.*]

A variety of other multiple-decision procedures have been considered, generally with the aim of mitigating oversimplification in the significance-testing approach. A common case is that in which one wants to choose the best of several populations, where "best" refers to largest expected value. [*A pioneering investigation along this line is Mosteller 1948. See also DECISION THEORY; SCREENING AND SELECTION.*]

**Hypotheses suggested by data.** In the course of examining a body of data, the analyst may find that the data themselves suggest one or more kinds of structure, and he may decide to carry out tests of corresponding null hypotheses from the very data that suggested these hypotheses. One difficulty here is statistical dependence (described above) between the tests, but there is a second difficulty, one that appears even if there is only a single significance test. In a sense, this difficulty is just a general form of the *post hoc* one-sided choice problem.

Almost any set of data, of even moderate size and complexity, will show anomalies of some kind when examined carefully, even if the underlying probabilistic structure is wholly random—that is, even if the observations stem from random variables that are independent and identically distributed. By looking carefully enough at random data, one can generally find some anomaly—for example, clustering, runs, cycles—that gives statistical significance at customary levels although no real effect



is present. The explanation is that although any particular kind of anomaly will occur in random data with .05 statistical significance just 5 times out of 100, so many kinds of anomalies are possible that at least one will very frequently appear. Thus, use of the same data both to suggest and to test hypotheses is likely to generate misleading statistical significance.

Most sets of real data, however, are not completely random, and one does want to explore them, to form hypotheses and to test these same hypotheses. One can sometimes use part of the data for forming hypotheses and the remainder of the data to examine the hypotheses. This is, however, not always feasible, especially if the data are sparse. Alternatively, if further data of the same kind are available, then one can use the earlier data to generate hypotheses and the later data to examine them. Again, this approach is not always possible.

As an example, consider one of the earliest instances of significance testing, by Daniel Bernoulli and his son John in 1734, as described by Todhunter ([1865] 1949, secs. 394–397). Astronomers had noticed that the orbital planes of the planets are all close together. Is this closeness more than one might expect if the orbital planes were determined randomly? If so, then presumably there is a physical reason for the closeness. The Bernoullis first attempted to make precise what is meant by randomness; in modern terminology this would be specification of randomness as a null hypothesis. They then computed the probability under randomness that the orbital planes would be as close together as, or closer than, the observed planes. This corresponds to deciding on a test statistic and computing a one-tail *P*-value. The resulting probability was very small and the existence of a physical cause strongly suggested. [For biographical material on the Bernoullis, see BERNOULLI FAMILY.]

Todhunter, in his *History*, described other early instances of significance testing, the earliest in 1710 by John Arbuthnot on the human sex ratio at birth (the relevant sections in Todhunter [1865] 1949 are 343–348, 617–622, 888, 915, and 987). Significance testing very much like that of the Bernoullis continues; a geophysical example, relating to the surprising degree of land–water antipodality on the earth's surface, has appeared recently (Harrison 1966). The whole topic received a detailed discussion by Keynes ([1921] 1952, chapter 25). Up-to-date descriptions of astronomical theories about the orbital planes of the planets have been given by Struve and Zeberg (1962, chapter 9).

Quite aside from the issue of whether it is reasonable to apply probabilistic models to such unique

objects as planetary orbits (or to the 50 states, as in an earlier example), there remains the difficulty that a surprising concatenation was noted in the data, and a null hypothesis and test statistic fashioned around that concatenation were considered. Here there is little or no opportunity to obtain further data (although more planets were discovered after the Bernoullis and, in principle, planetary systems other than our sun's may someday be observable). Yet, in a way, the procedure seems quite reasonable as a means of measuring the surprisingness of the observed closeness of orbital planes. I know of no satisfactory resolution of the methodological difficulties here, except for the banal moral that when testing hypotheses suggested during the exploration of data, one should be particularly cautious about coming to conclusions. (In the Bayesian approach to statistics, the problem described here does not arise, although many statisticians feel that fresh problems are introduced [see BAYESIAN INFERENCE].)

An honest attempt at answering the questions, What else would have surprised me? Was I bound to be surprised? is well worthwhile. For example, to have observed planetary orbits symmetrically arranged in such a way that their perpendiculars nearly formed some rays of a three-dimensional asterisk would also have been surprising and might well be allowed for.

**Difficulties in determining reference set.** Some statisticians have been concerned about difficulties in determining the proper underlying reference probabilities for sensibly calculating a *P*-value. For example, if sample size is random, even with a known distribution, should one compute the *P*-value as if the realized sample size were known all along, or should one do something else? Sample size can indeed be quite random in some cases; for example, an experimenter may decide to deal with all cases of a rare mental disorder that come to a particular hospital during a two-month period. [Three papers dealing with this kind of problem are Barnard 1947; Cox 1958; and Anscombe 1963. See also LIKELIHOOD.]

Random sample sizes occur naturally in that part of statistics called sequential analysis. Many sequential significance testing procedures have been carefully analyzed, although problems of determining reference sets nonetheless continue to exist. [See SEQUENTIAL ANALYSIS.]

**Optional stopping.** Closely related to the discussion of the preceding section is the problem of optional stopping. Suppose that an experimenter with extensive resources and a tendentious cast of mind undertakes a sequence of observations and

from time to time carries out a significance test based on the observations at hand. For the usual models and tests, he will, sooner or later, reach statistical significance at any preassigned level, even if the null hypothesis is true. He might then stop taking observations and proclaim the statistical significance as if he had decided the sample size in advance. (The mathematical background of optional stopping is described in Robbins 1952.)

For the standard approach to significance testing, such optional stopping is as misleading and reprehensible as the suppression of unwanted observations. Even for an honest experimenter, if the sampling procedure is not firmly established in advance, a desire to have things turn out one way or another may unconsciously influence decisions about when to stop sampling. If the sampling procedure is firmly established in advance, then, at least in principle, characteristics of the significance test can be computed in advance; this is an important part of sequential analysis.

Optional stopping is, of course, relevant to modes of statistical analysis other than significance testing. It poses no problem for approaches to statistics that turn only on the observed likelihood, but many statisticians feel that these approaches are subject to other difficulties that are at least equally serious. [See BAYESIAN INFERENCE; LIKELIHOOD.]

**Simplicity and utility of hypotheses.** It is usually the case that a set of data will be more nearly in accord with a complicated hypothesis than with a simpler hypothesis that is a special case of the complicated one. For example, if the complicated hypothesis has several unspecified parameters whereas the simpler one specializes by taking some of the parameters at fixed values, a set of data will nearly always be better fit by the more complicated hypothesis than by the simpler one just because there are more parameters available for fitting: a point is usually farther away from a given line than from a given plane that includes the line; in the polynomial regression context, a linear regression function will almost never fit as well as a quadratic, a quadratic as well as a cubic, and so on.

Yet one often prefers a simpler hypothesis to a better-fitting more complicated one. This preference, which undoubtedly has deep psychological roots, poses a perennial problem for the philosophy of science. One way in which the problem is reflected in significance testing is in the traditional use of small significance levels. The null hypothesis is usually simpler than the alternatives, but one may be unwilling to abandon the null hypothesis unless the evidence against it is strong.

Hypotheses may be intrinsically comparable in ways other than simplicity. For example, one hypothesis may be more useful than another because it is more closely related to accepted hypotheses for related, but different, kinds of observations.

The theory of significance testing, however, takes no explicit account of the simplicity of hypotheses or of other aspects of their utility. A few steps have been made toward incorporating such considerations into statistical theory (see Anderson 1962), but the problem remains open.

**Importance of significance testing.** Significance testing is an important part of statistical theory and practice, but it is only one part, and there are other important ones. Because of the relative simplicity of its structure, significance testing has been over-emphasized in some presentations of statistics, and as a result some students come mistakenly to feel that statistics is little else than significance testing.

**Other approaches to significance testing.** This article has been limited to the customary approach to significance testing based on the frequency concept of probability. For other concepts of probability, procedures analogous to significance testing have been considered. [See BAYESIAN INFERENCE. An extensive discussion is given in Edwards et al. 1963.] Anscombe (1963) has argued for a concept of significance testing in which only the null hypothesis, not the alternatives, plays a role.

WILLIAM H. KRUSKAL

[See also HYPOTHESIS TESTING.]

#### BIBLIOGRAPHY

- ANDERSON, T. W. 1962 The Choice of the Degree of a Polynomial Regression as a Multiple Decision Problem. *Annals of Mathematical Statistics* 33:255-265.
- ANScombe, F. J. 1963 Tests of Goodness of Fit. *Journal of the Royal Statistical Society Series B* 25:81-94.
- BAKAN, DAVID 1966 The Test of Significance in Psychological Research. *Psychological Bulletin* 66:423-437.
- BANCROFT, T. A. 1964 Analysis and Inference for Incompletely Specified Models Involving the Use of Preliminary Test(s) of Significance. *Biometrics* 20:427-442.
- BARNARD, G. A. 1947 The Meaning of a Significance Level. *Biometrika* 34:179-182.
- BORING, EDWIN G. 1919 Mathematical vs. Scientific Significance. *Psychological Bulletin* 15:335-338.
- COX, DAVID R. 1958 Some Problems Connected With Statistical Inference. *Annals of Mathematical Statistics* 29:357-372.
- EDWARDS, WARD; LINDMAN, HAROLD; and SAVAGE, LEONARD J. 1963 Bayesian Statistical Inference for Psychological Research. *Psychological Review* 70:193-242.
- GOOD, I. J. 1956 The Surprise Index for the Multivariate Normal Distribution. *Annals of Mathematical Statistics* 27:1130-1135.



- GOOD, I. J. 1958 Significance Tests in Parallel and in Series. *Journal of the American Statistical Association* 53:799-813.
- HARRISON, CHRISTOPHER G. A. 1966 Antipodal Location of Continents and Oceans. *Science* 153:1246-1248.
- HIRSCHI, TRAVIS; and SELVIN, HANAN C. 1967 *Methods in Delinquency Research*. New York: Free Press. → See especially Chapter 13, "Statistical Inference."
- HODGES, J. L. JR.; and LEHMANN, E. L. 1954 Testing the Approximate Validity of Statistical Hypotheses. *Journal of the Royal Statistical Society Series B* 16: 261-268.
- KAISER, HENRY F. 1960 Directional Statistical Decisions. *Psychological Review* 67:160-167.
- KEYNES, JOHN MAYNARD (1921) 1952 *A Treatise on Probability*. London: Macmillan. → A paperback edition was published in 1962 by Harper.
- KITAGAWA, TOSIO 1963 Estimation After Preliminary Tests of Significance. *University of California Publications in Statistics* 3:147-186.
- MOSES, LINCOLN E. 1956 Statistical Theory and Research Design. *Annual Review of Psychology* 7:233-258.
- MOSTELLER, FREDERICK 1948 A  $k$ -sample Slippage Test for an Extreme Population. *Annals of Mathematical Statistics* 19:58-65.
- MOSTELLER, FREDERICK; and BUSH, ROBERT R. (1954) 1959 Selected Quantitative Techniques. Volume 1, pages 289-334 In Gardner Lindzey (editor), *Handbook of Social Psychology*. Cambridge, Mass.: Addison-Wesley.
- NEYMAN, JERZY 1958 The Use of the Concept of Power in Agricultural Experimentation. *Journal of the Indian Society of Agricultural Statistics* 9, no. 1:9-17.
- ROBBINS, HERBERT 1952 Some Aspects of the Sequential Design of Experiments. *American Mathematical Society, Bulletin* 58:527-535.
- SAVAGE, LEONARD J. 1954 *The Foundations of Statistics*. New York: Wiley.
- STERLING, THEODORE D. 1959 Publication Decisions and Their Possible Effects on Inferences Drawn From Tests of Significance—or Vice Versa. *Journal of the American Statistical Association* 54:30-34.
- STRUVE, OTTO; and ZEBERG, VELTA 1962 *Astronomy of the 20th Century*. New York: Macmillan.
- TODHUNTER, ISAAC (1865) 1949 *A History of the Mathematical Theory of Probability From the Time of Pascal to That of Laplace*. New York: Chelsea.
- TULLOCK, GORDON 1959 Publication Decisions and Tests of Significance: A Comment. *Journal of the American Statistical Association* 54:593 only.
- WEAVER, WARREN 1948 Probability, Rarity, Interest, and Surprise. *Scientific Monthly* 67:390-392.
- WITTENBORN, J. R. 1952 Critique of Small Sample Statistical Methods in Clinical Psychology. *Journal of Clinical Psychology* 8:34-37.
- WOOFER, T. J. JR. 1933 Common Errors in Sampling. *Social Forces* 11:521-525.

## SIMIAND, FRANÇOIS

François Simiand (1873-1935), French economist and economic historian, was born at Gières (Isère) to a large family of modest means. Like most members of his family, he became a teacher. At 20, he entered the École Normale Supérieure in

Paris, preparing first for his teacher's license and then for his *agrégation* in philosophy. He passed the *agrégation* in 1896 and was then admitted to the Fondation Thiers, newly opened to researchers. At that time Simiand became interested in sociology and attended the courses taught by Émile Durkheim and Lucien Lévy-Bruhl. He began work on a dissertation on miners' wages in France, and thenceforth his sociological orientation distinguished much of his work from that of his economist contemporaries.

When Simiand was at the École Normale, Parisian students were very much stirred up by the Dreyfus affair. He and his fellow students had several opportunities to hear lectures by the socialists Lucien Herr and Jean Jaurès, and like many other Normaliens Simiand came to have a lasting sympathy for socialism. With Herr, Charles Péguy, and Léon Blum he later founded a bookselling and publishing cooperative, the Librairie Georges Bellais, named for one of their friends. Simiand's principal task was to edit a popular series, the *Bibliothèque socialiste*, and to head a technical program for training personnel for the cooperative. Since he had to earn a living at the same time that he was joining in the battle for socialism, he took a position as librarian at the Ministry of Labor, established in 1906, where he performed the enormous task of providing the documentary resources necessary for the new bureau.

At this time Simiand wrote methodological discussions that were violently critical of the historians, particularly of Charles Seignobos. The best example is an article entitled "Méthode historique et science sociale," which appeared in the *Revue de synthèse historique* in 1903. Simiand rejected Seignobos' conception of history as a mere collection of political facts isolated from their social and economic context. Even such faithful friends as Paul Mantoux, however, could not accept Simiand's destructive conclusions. Simiand was also at work on his thesis, *Le salaire des ouvriers des mines de charbon en France*, which he defended in 1904 and published in a revised version in 1907. It was the forerunner of his later research on wages.

With the war of 1914 Simiand began a period of government service. His socialist connections led to his appointment as aide to Albert Thomas, who became undersecretary of state for artillery and munitions in 1915. Simiand continued in government service until 1920. He worked strenuously to organize war production; later he saw to it that the transition from the German to the French regime in Alsace-Lorraine was carried out in such a way that the people of those provinces retained the social gains they had made before 1918.

In 1920 Simiand resigned his administrative functions and from then on devoted himself entirely to study and teaching. In 1910 he had been called to the École Pratique des Hautes Études as lecturer in the history of economic theory. After 1920 he resumed teaching there, and in 1924 he became chairman of the department of history and economic statistics. At the same time he also taught at the Conservatoire National des Arts et Métiers, as professor of labor organization and labor unions from 1919 to 1923 and as professor of political economy from 1923 to 1933. In 1932 he succeeded Georges Renard as professor of labor history at the Collège de France. Simiand's health, already impaired by his work during the war, could not stand the intense pace of his activity. He died at Saint-Raphael (Var) in 1935.

This final period, from about 1930 to 1935, was intellectually Simiand's most fruitful one, and his publications were numerous. He published the lectures he gave at the Conservatoire, which clearly show the development of his economic thinking. He also produced works inspired by the world crisis—a set of lectures given at the University of Brussels on long-term economic fluctuations (1932a) and a study of economic development in the United States, in which economic analysis is related to historical factors (1934).

Most important, Simiand completed the master-work to which he had devoted so much of his life, his research on price movements and wages (1932b). He prefaced the work by a long methodological introduction, explaining his particular interest in the study of wages. Wages, he asserted, are a social reality, and it was by the use of economics that he hoped to grasp social realities of this kind. Furthermore, wages are an empirical entity: they can be dealt with statistically; they correspond to a real value; they are part of an economic and social complex. Accordingly, wages appear to be made to order as a subject for the kind of scientific investigation advocated by Simiand, namely one that applies the methods of the natural sciences to the social sciences. As he saw the study, the methodological procedures gradually emerged from the study itself and were not established by a conceptual system nor did they follow preconceived ideas. Simiand believed that he had thus found a new method for studying economic facts, but his ideas found no followers, and he remained an isolated figure.

CLAUDE FOHLEN

[Other relevant material may be found in the biographies of BLOCH; DURKHEIM; HALBWACHS; LEVASSEUR; LÉVY-BRUHL.]

#### WORKS BY SIMIAND

- 1903 *Méthode historique et science sociale. Revue de synthèse historique* 6:1–22, 129–157.  
 (1904) 1907 *Le salaire des ouvriers des mines de charbon en France: Contributions à la théorie économique du salaire*. Paris: Cornély. → An expanded version of Simiand's thesis.  
 1912 *La méthode positive en science économique*. Paris: Alcan.  
 1922 *Statistique et expérience: Remarques de méthode*. Paris: Rivière.  
 1930a *Cours d'économie politique professé en 1928–1929*. Paris: Domat-Montchrestien.  
 1930b *Cours d'économie politique professé en 1929–1930*. Paris: Loviton.  
 (1931) 1932 *Le salaire, l'évolution sociale et la monnaie; essai de théorie expérimentale du salaire: Introduction et étude globale*. 3 vols. Paris: Alcan.  
 1932a *Les fluctuations économiques à longue période et la crise mondiale*. Paris: Alcan.  
 1932b *Recherches anciennes et nouvelles sur le mouvement général des prix du XVI<sup>e</sup> au XIX<sup>e</sup> siècle*. Paris: Domat-Montchrestien.  
 1932c *Cours d'économie politique professé en 1930–1931*. Paris: Loviton.  
 1934 *Inflation et stabilisation alternées: Le développement économique des États-Unis (des origines coloniales au temps présent)*. Paris: Domat-Montchrestien.

#### SUPPLEMENTARY BIBLIOGRAPHY

- BOUGLÉ, CÉLESTIN 1936 *La méthodologie de François Simiand et la sociologie. Annales sociologiques* Series A 2:5–28.  
 DAMALAS, BASILE 1943 *L'oeuvre scientifique de François Simiand*. Paris: Presses Universitaires de France.  
 → See especially the bibliography on pages 267–270.  
 HARSIN, PAUL 1935 *François Simiand: 1873–1935. Société Scientifique de Bruxelles, Annales* Series D 55: 39–68.  
 LAZARD, MAX 1936 *François Simiand, 1873–1935: L'homme, l'oeuvre*. Paris: Domat-Montchrestien.  
 MORAZÉ, CHARLES 1942 *La leçon d'un échec: Essai sur la méthode de François Simiand. Mélanges d'histoire sociale* 1:5–24; 2:22–44.

#### SIMMEL, GEORG

Georg Simmel (1858–1918), German philosopher and sociologist, is still a controversial figure. While some hail him as the founder of modern sociology, others see in him only a brilliant stylist who made no original contribution and failed to develop a systematic theory. It is true that Simmel was no system builder and was often preoccupied with defining approaches—philosophical, historical, or sociological—which he illustrated with his substantive analyses. But these analyses, subtle and full of surprising insights, still delight and stimulate readers, while many of the more systematic works of his contemporaries are no longer read.

**Life and career.** Simmel's life was externally uneventful. He spent most of it in Berlin, his native



city; he traveled little; and he experienced neither true poverty nor wealth, neither unanimous rejection nor great success. His predominant preoccupations were intellectual. He was not immediately concerned with the acute social and political problems of his age, and only once did a political event, the outbreak of World War I, move him to deep and emotional concern. His intellectuality was passionate, and he had a strong disdain for everything crude and vulgar, reflecting a kind of spiritual aristocracy.

Simmel's father was a Jewish businessman who had become a Roman Catholic; his mother was of Jewish origin but a Lutheran. Simmel, baptized as a Lutheran, later renounced his church membership but retained a philosophical interest in religion. While he was still in school, his father died. His guardian later left him a small fortune, which enabled Simmel to live as a scholar even when he was without a salaried academic position. He studied history, folk psychology, the history of art, and philosophy at the University of Berlin, where he received his doctorate in 1881. His dissertation on the philosophy of Kant gained a prize; the thesis he wrote for his subsequent habilitation also dealt with Kant. He started to lecture as a *Privatdozent* in 1885 and almost immediately attracted large and enthusiastic audiences. At first he lectured on philosophy and ethics, but soon he began to emphasize certain sociological aspects of his topics and later devoted whole courses to this new subject, which at that time was not being taught by anyone else at his university. However, sociology accounted for only about half of his course offerings.

In 1890 Simmel married. His wife, Gertrud, with whom he had one son, was also intellectually inclined and a writer on philosophy. Their home became a meeting place for members of Berlin's intellectual elite. But in his academic career Simmel was long frustrated, apparently because of his Jewish origin, his nonprofessorial brilliance, and his intellectual attitude, which some felt to be destructive. It was 15 years before he even became titular professor, and all attempts to achieve a full professorship failed until, at the age of 56, he was finally called to a chair at the University of Strassburg in 1914. He died there of cancer in the fall of 1918.

**Influence.** Simmel may be remembered today primarily by sociologists, but for him sociology was only one of many interests, and his productive period in this field lasted barely more than one decade. His earliest and his latest works were philosophical. He had a lifelong interest in art, admired and wrote about Rodin, received the young

Rilke in his home, published a book about Goethe, and was a friend of Stefan George. Simmel's bibliography (Gassen & Landmann 1958) lists 31 books (some very short ones), 256 articles, essays, etc. (many published in nonscholarly journals), and about 100 translations of his works into other languages.

Thus, Simmel was well known during his lifetime. However, he did not found a "school" in any of his fields of work. In philosophy his reputation did not long survive his death; in the 1920s his work was already overshadowed by existential philosophy. In sociology, Tönnies praised him during his lifetime, Durkheim published one of his essays in *L'année sociologique*, and Max Weber, whom he joined in founding the German Sociological Society in 1910, esteemed him and in his own later work utilized some of Simmel's concepts and theses. However, both Durkheim and Weber soon surpassed Simmel in influence and renown, and later only Leopold von Wiese followed the path outlined by Simmel, although even von Wiese coupled his explicit indebtedness with criticism.

In the United States, Simmel first became known through early translations, many of them by Albion W. Small, and through the efforts of Park and Burgess (1921) and Spykman (1925). But the impact was not deep. After a period of oblivion, the efforts of Kurt H. Wolff (see *The Sociology of Georg Simmel*; Simmel 1908a; Wolff 1959) and Rudolph H. Weingartner (1962) have revived some interest in Simmel, but with a few exceptions, such as Coser's book on conflict (1956) and the work of several authors on the three-person group (the triad), Simmel's analyses have not been explicitly followed up by modern sociologists.

Simmel himself foresaw this fate when he wrote in his diary that he would leave no spiritual heirs because his heritage would be dispersed and made use of as money is used. And indeed, much of his sociological thought has become common coin without remaining associated with his name.

**Intellectual ancestors.** Simmel was sensitive to the intellectual currents of his time and reflected most of them in his work without identifying himself exclusively with any one of them. His methodology is fashioned according to Kantian principles, but with modifications. He was influenced by Hegel, Schopenhauer, and Nietzsche, and his philosophy of life—foreshadowing modern existentialism—has noticeable affinities with the work of his friend Husserl and of Bergson. The influence of Dilthey can be found in both Simmel's sociology and his cultural philosophy, notably in his views on differentiation, interaction (*Wechselwirkung*), and

the individual. Spencer, Comte, and Marx were other influences on Simmel's sociology, but he rejected contemporary positivism and social Darwinism, and he was definitely not a Marxist. Thus, the work of Simmel escapes easy classification.

**Philosophical position.** For Simmel, philosophy was not a discipline ranged alongside other disciplines. Philosophy both precedes and follows all special disciplines—as epistemology, on the one hand, and as philosophy of the particular object, on the other. Philosophy, then, is not constituted by and confined to a particular subject matter; it is a distinct mode of treating any given subject matter, characterized by receptiveness to the totality of being and at the same time expressive of a fundamental attitude or world orientation on the part of the philosophizing person. Simmel demonstrated this view by writing on the philosophy of such different objects as money, culture, the sexes, religion, and art.

Epistemologically, Simmel held, in the manner of Kant, that all experience of contents is shaped by a priori categories. This view is the basis for his distinction between form and content, which later also became the methodological principle for his sociology. The distinction is clearly analytical. Form is, first of all, a basic organizing principle of perception or modality of experience, which imparts structure to that which in its immediate unity is structureless. Not simply philosophy itself, but also religion and science, are modalities of experience that may be applied alternatively to the same content.

Simmel extended this view to history. History is not the exact reproduction of all that has happened and is happening, but a coherent representation of the boundless manifold of human life and its products, created by selective perception and ordering. At first Simmel held this to be true not only for historiography but also for the conscious experience of history; later he modified this view somewhat and found a certain intrinsic coherence in history itself as a process of life. But Simmel did not stipulate any specific selective principle as adequate to the historical mode of perception, and it was this relativism that Troeltsch criticized as excessive historicism. Simmel never worked as a historian in the usual sense, although he had a widely ramified knowledge of history and often drew on it for illustrations, especially in his sociological works—including as sociological his *Philosophie des Geldes* (1900a).

Simmel's philosophies of life and of culture may be considered together as one coherent, substantive theory. Life as a force or principle was for him a

process of unbounded continuity, an uninterrupted flowing, a creative movement: life is more-life, as he put it. As a principle, life is the opposite of form and alien to it. But as a concrete process, life can become real only in forms. The individual, the concrete subject of life, is himself a bounded creature, and individual experience gives form to contents, thereby creating the objects of external reality. This process of objectification has two aspects: the perception of objects as external and the production of objects of material and spiritual culture (e.g., law, systems of knowledge, art, religion). Here Simmel talked of life as more-than-life, life becoming real and expressing itself in the creation of objects. These objects he also called "forms," a duplication of terms that may easily cause confusion.

As a continuous and creative process, life cannot be contained in finished and static forms (or objects). It must leave and destroy them, but only to crystallize again in new forms. This is the basic antinomy of life, which becomes the tragedy of culture when attention is turned to the fate of the individual in this process. The individual produces objects of culture to maintain his life and develop his potentialities. To this end he must not only utilize the sum total of human products (or objective spirit), but he must also interiorize them and reintegrate them into the stream of his life. But this reintegration of subject and object remains unachievable. The elements of objective spirit, as finished forms dissociated from the stream of life, possess their own dynamic. They become autonomous and develop according to their inner logic, no longer means but ends in themselves. Thus, if man first wanted to know in order to live, some men will later live only to accumulate more knowledge. In the end, man is dominated and enslaved by his own products.

The parallel to the dialectic theory of alienation is obvious. But if for Marx alienation can be resolved in a future society, for Simmel the contradiction flowing from the antinomy of life is eternal. Naturally, Marxist critics jumped on this. Thus, in a brilliant critique, Lukács (1953) accused Simmel of promoting a cynical and complacent adjustment to the *status quo* by interpreting as eternal a condition of the individual that is characteristic only of imperialism.

Simmel's conception of the process of culture is not completely ahistorical. As he noted especially in the *Philosophie des Geldes* (1900a), social differentiation and the attendant specialization of human activities greatly increase the production of cultural objectifications. Thus, at the same time



that specialization makes the full development of personality ever more difficult, there is a steadily growing amount of objective culture to be interiorized if subjective (or personal) culture is to flourish. Accordingly, the discrepancy between objective and subjective culture appears largest in contemporary society, where it has given rise to a struggle not only against obsolete forms, but against form as a principle.

The development of individual personality is a concern that links Simmel's philosophy with his sociology. In the latter, he came to a more positive conclusion. While the full realization of individual personality is impeded by the tragedy of culture, it is also promoted by social differentiation and the expansion of social groups, and potentially even by the growth of a money economy.

**The process of social development.** Simmel's theory of social development, although presented in his sociological writings, can be considered apart from his "pure" or "formal" sociology. In viewing this process, Simmel used two distinct points of reference, or approaches. The first, linked to his view of the process of culture, considers social differentiation and group expansion chiefly from the point of view of the effects on individualization.

Social differentiation and the quantitative expansion of groups are seen to be closely interrelated. Small groups are more vulnerable and cannot make full use of the energy-saving principle of division of labor; hence their preservation requires greater involvement and effort on the part of each member. With the expansion of groups and increasing differentiation, a lesser involvement is sufficient—and is all that can be demanded, since the consensus in larger groups tends to become segmental. Also, with group expansion the relations between members become impersonal, guided by an order of abstract rules. Thus only a limited and clearly defined contribution comes to be expected of each member, and this results in greater personal freedom. At the same time individuality has a chance to increase: as groups form on the basis of segmental interest and partial involvement, each person tends to belong to a larger number of different groups, and this has an individualizing effect.

The development of a money economy contributes to individualization both directly and indirectly, by stimulating the growth of large and purposive groups. Money is the basis of impersonal social relations and permits the substitution of financial for personal obligations. Of all types of obligation, a merely financial one is most compat-

ible with a maximum of personal freedom—although this is, of course, freedom only in the negative sense, or potential freedom, which must be used for personal development. As money makes the individual independent of particular persons and groups, it stimulates the formation of associations for limited and specific purposes and hence contributes to the growth and multiplication of groups with the attendant individualizing effect.

Simmel saw a close relationship between the existence of a money economy, rationality, and the growth of natural science. All three are based on the principle of calculation. Rational (i.e., purposive and calculating) action presupposes knowledge of causal relations and hence motivates the search for such knowledge. As a consequence of such pervasive rationalization, emotionalism decreases; this, on the one hand, means a shallower emotional life but, on the other hand, facilitates mutual understanding and conciliation.

Simmel's second approach in considering the process of group expansion and social development would today be called functionalist. It derives from the influence of Spencerian evolutionism, which marked Simmel's view on social differentiation as put forth in his first sociological publication (1890). There he stated that the development of organisms is directed by a tendency toward greater efficiency, or as he put it, toward a relative saving of energy. This tendency is, according to Simmel, the characteristic, and positive, consequence of social differentiation, and he took great pains to demonstrate this principle in the growth of various forms of higher social organization—not very convincingly for the modern reader. In fact, Simmel dropped this evolutionary thesis later and did not repeat it in his other sociological works. But he retained the basic principle implied in this thesis—the explanation of certain social forms, processes, and changes by their usefulness.

Simmel's functionalism pervades all his sociological work, but it is perhaps most prominent in his treatment of the self-preservation of social groups, where he discussed the conditions for the continuous existence of a group in spite of a constant turnover of members—among these conditions are functional differentiation and formation of specialized organs, symbols, and norms, and formalization or impersonality in the definition of duties and obligations. However, here at least Simmel escaped the danger of organismic reification, which threatens any attempt to explain the rise or persistence of something by its usefulness to the group or social system. In his book on social differen-

tiation (1890) he had stated explicitly that the observable tendencies toward group maintenance—such as group defense mechanisms—are not independent forces inherent in the group as such. Rather, they result from the self-interest of the members in preserving the group, because group solidarity is a necessity for individual life, especially in situations of conflict and strife. By virtue of social-psychological mechanisms, which Simmel only sometimes treated explicitly, the solidarity thus created tends to survive its immediate causes, giving existing groups a greater stability.

**Relationism and the concept of society.** One of Simmel's basic tenets is that all things are to be considered as interdependent, or as functions of each other. This view, referred to as his functionalism, relativism, or, most appropriately, relationism, should be clearly distinguished from his specifically sociological functionalism, which, by contrast, went largely unnoticed until recently. Simmel's relationism derives from his basic philosophical position that reality is essentially movement, continuity, process, and that only the human intellect, fashioned to serve as an instrument for action and not for gaining knowledge for its own sake, tends to perceive reality in terms of structures and substances—solid, fixed phenomena that are amenable to classification, ordering, and calculation. But what is thus commonly regarded as a substance is often essentially a process, a function, an interdependence.

This view is most pronounced in Simmel's treatment of money and of society. For Simmel money is not in itself a valuable substance and therefore exchangeable for objects of all sorts; it is, on the contrary, essentially the action of exchange in a concrete form, the material expression of the exchange relationship, a social function become a substance [see INTERACTION, *article on SOCIAL INTERACTION*].

Simmel similarly rejected the view that society is a substance, an organism, an irreducible real entity. Society is nothing but the sum total of the interactions and interdependencies (the German term *Wechselwirkung* implies both) between individuals—whose unity in turn is constituted only by the interaction between parts. Understandably, this view of society met with criticism from authors such as Othmar Spann, who resented this dissolution of the social organism and charged Simmel with psychologizing sociology. In fact, Simmel's explicit reference in his definition of society to Kant's treatment of nature may be misleading. But society was for him not merely a psychic phenom-

enon, a phenomenon of perception, because its constituents—interactions and interdependencies—are real enough. Only the conception of society as an entity is a matter of perception.

**The nature of sociology.** Simmel's lasting attractiveness as a sociological author may derive from the subtlety of his substantive analyses, but his most important contribution is, nevertheless, usually considered to be his methodology, which gave an independent status to sociology as a special discipline. Simmel rejected sociology as a comprehensive study of all social facts, since as such it would be only a label for numerous special disciplines treating different aspects of the same general object, human life and its products. But if sociology has no concrete object of its own, it nevertheless gains a special object of cognition by isolating analytically the aspect of interaction from the whole of social reality. Interactions are to be analyzed not only where they have crystallized into relatively permanent structures or social institutions, but also on the level of the more transitory interpersonal relations. This emphasis on the sub-institutional, or microsociological, was novel for his time.

Simmel was also impatient with the confusion between sociology and the philosophy of history or social philosophy, and he wanted sociology to keep close to observable phenomena, leaving all metaphysical interpretations that transcend objective knowledge to what he later (1917) called "philosophical sociology." However, Simmel did not conceive of sociology in a positivistic fashion, as a science of laws. He had early rejected the use of natural science as a model for the disciplines dealing with man and his creations, and he did not side with the positivists in the hotly debated issue of the idiographic as against the nomothetic character of social science. Rather, he attempted to play down any such distinction by pointing out that both viewpoints are alternative modes of perception. But the search for social laws never appeared promising to him. As early as 1890 he wrote that since a given event may have several and alternative causes, so that it can be produced by varying constellations of factors, it is usually impossible to formulate rules about a lawful association of isolated variables. When he himself described some specific regularity in his substantive analyses he repeatedly pointed out that this was not to be taken as a "law," but only as a descriptive generalization, leaving quite open which causes might produce it under varying conditions. Simmel expressly denied the existence of laws of social



development or of history in general and once even suggested that what we now know as natural laws may at some later time be revealed to be superstitions. Lukács (1953) has charged that this "struggle against law and causality" was the self-defense of imperialistic philosophy against dialectical materialism and that with it Simmel undermined objective scientific knowledge.

Of course, Simmel did not reject the possibility of causal explanation in the realm of human life and its products. But such explanation was for him essentially psychological in nature. Ultimately, the genesis of specific social forms and cultural objectifications must be derived psychologically. In his substantive analyses, Simmel again and again offered psychological explanations for the phenomena he described, both in terms of the conscious motives, interests, or goals of the individuals involved, and in terms of unconscious psychic mechanisms determining the reactions of these individuals. But he did not draw on any systematic psychological theory; rather, his often strikingly sensitive and perceptive explanations seem the result of introspection and conjecture.

**Formal sociology.** In 1890 Simmel defined the subject matter of sociology as consisting of interactions and interrelations, but he did not remain content with this definition. Later, going back to the basic distinction made in his philosophy, he demanded that the forms of interaction be distinguished from their contents and held that only the former constitute sociology's special object of cognition. Still later (1917), this seems to have appeared too restrictive to him, and he conceived of a "general sociology" in addition to formal sociology; but he failed to make its approach clear.

The contents of interaction are its motive powers—individual drives, purposes, interests, and so forth. The forms of interaction, or social forms, are not structured real entities (analogous to the forms of culture) but abstract, analytical aspects of social reality. They might best be conceived of as basic structural configurations or structuring principles. Simmel found a justification for this distinction in the seeming independence of forms and contents: the same structural principle can be observed in groups or relations characterized by widely divergent contents, while the same content (interest, purpose) may be found expressed in different forms of social organization.

Simmel's attempt to explain the concept of social form with a misleading analogy between sociology and geometry and his admitted failure to make his methodological principle quite explicit did much to provoke criticism of his "formal sociol-

ogy." Such criticism was especially directed against the form-content distinction, on the grounds that in social reality the structural details of any concrete group and its particular syndrome of "contents" are inseparably linked. This is true and, incidentally, is quite obvious from Simmel's own analyses. But Simmel expressly distinguished the forms of sociation from their embodiment in special social units. Conceived of as very general, abstract structural principles, or basic modes of interrelationship among social elements, the forms of sociation can indeed be isolated analytically. Structural principles that Simmel himself named are the relation of superordination-subordination; the relation of antagonism (conflict); the division of labor, or relation of functional interdependence; the ingroup-outgroup relation and the related principle of party formation; the principle of representation; the principles of spatial and temporal structuring; and the quantitative dimension. Other structural principles are implied in Simmel's writings, such as the dependence-autonomy dimension, which plays such a crucial role in his analysis of group membership and individuality.

The specific task of sociology was for Simmel not the causal explanation of these social forms, but the analysis of their objective meaning. This distinction is analogous to one commonly made in the analysis of forms of culture, such as objects of art and systems of ideas: they possess an *objective meaning*, which is something distinct from their psychologically explained genesis. The objective meaning (the analogy does not fit completely) of social forms rested for Simmel in their essential characteristics and range of empirical variation, on the one hand, and in their consequences, on the other. His *Soziologie* (1908b) contains several chapters devoted to the analysis of selected structural principles in these terms. In other chapters he started by focusing on a specific type of group (e.g., secret societies) or class of persons (e.g., the stranger, the poor man, the aristocrat) and proceeded to show how they are characterized and determined by a unique constellation of several of these structural principles.

Where Simmel dealt with the essential characteristics of some social form, his writing has a strongly phenomenological flavor. But the inclusion of the consequences or effects of social forms breaks this confinement and links his formal approach with his theory of social development, since he selected for analysis the consequences of given structural configurations either for individualization or for the preservation and functioning of groups. This procedure is obvious not only in the

topical chapters on these subjects, but also where he treated specific structural principles (e.g., conflict), special types of relationship, such as marriage, or special types of groups, such as secret societies. Thus his *Soziologie*, often criticized for being an unsystematic collection of substantive analyses only vaguely connected with his formal approach, really does possess a surprising—if partly implicit—internal coherence.

**Norms and individual behavior.** Simmel occasionally used the term "role" and recognized that social relations are defined by mutual obligations. Yet it would be a mistake to interpret his social forms as structures in the Parsonian sense, i.e., basically normative patterns, or to see the influence of social forms on individual behavior as that of role expectations. When Simmel analyzed individual behavior, he saw it primarily as a result of individual motives and of psychologically mediated reactions to the structure of the situation. He would not have denied that norms also play a part in determining individual behavior, but he paid little attention to them on this level of analysis. This may be due to his often paramount psychological interest and to the fact that he did not conceive of norms as external constraints or impositions. The latter view is made explicit in his *Lebensanschauung*, or philosophy (1918a), where he maintained that *Sollen*—a substantivation of the German verb meaning "shall" for which there is no exact translation—is a basic mode of life, a "categorical archphenomenon." Its specific contents, however, are to be derived psychologically and historically.

This last view seems connected with the fact that when Simmel argued sociologically, on the group level instead of on the individual level, he paid more explicit attention to norms. And here, the "historical" determination of the content of norms appears to be essentially of a functionalistic nature. In 1890 he had stated that what is "correct behavior" changes with the expansion of the group, and from several examples he used for illustration it becomes clear that the content of a norm, be it law or custom, depends on its social usefulness. Similar examples can be found in his later sociological works. But the whole principle is never explicitly stated as such, and the criterion of usefulness is not defined precisely; sometimes it seems to be goal achievement, sometimes group preservation, and sometimes it is the facilitation of a process of development.

**Ahistorical nature of formal sociology.** One of the major criticisms of Simmel's approach is directed at its ahistorical nature. Of course, Simmel

did not deny the historical character of society, which he saw as a continuous process, the present always being shaped by the accumulated past. But it seems that by selecting the forms of interaction for analysis, he defined as the essence of society that which is changeless and ahistorical, thus freezing history into motionless structures by his mode of perception.

That he conceived of the forms of sociation as abstract, general principles is undeniable. He used them to *describe* historical forms of social organization and certain general aspects of social development, but he did not use them to *explain* social phenomena as stages in a historical process. His focus on basic structural characteristics even made him occasionally describe as permanent some social form or relation that we would recognize as historically relative. This is true, for instance, of his description of sociability as a pure form of sociation, which might be roughly identified as an upper-bourgeois culture pattern, or of his analysis of the characteristics of man and woman, which is in effect largely a generalization of culturally relative sex roles.

The question is whether these examples merely point to occasional errors in Simmel's search for truly general ahistorical structuring principles or whether he failed—whether he *had* to fail—in this effort throughout. For it might be argued that since social reality is always historical in nature, nothing but historical forms can be abstracted from it. If this were true, even Simmel's most general structural principles, like antagonism, superordination, and so forth, would be specific to a certain phase of history and inapplicable in others. Simmel did not think so, which means that as a matter of principle he thought it possible to abstract ahistorical concepts or analytical dimensions from social reality. While the majority of contemporary sociologists would tend to share his view, it is this methodological position which, in the last analysis, defines Simmel as ahistorical in the eyes of modern dialectic theory.

RENATE MAYNTZ

[For the historical context of Simmel's work, see ALIENATION; CONFLICT, article on SOCIAL ASPECTS; HISTORY, article on THE PHILOSOPHY OF HISTORY; INTERACTION, article on SOCIAL INTERACTION; SOCIETY; and the biographies of DILTHEY; HUSSERL; TROELTSCH. For discussion of the subsequent development of his ideas, see COALITIONS, article on COALITION FORMATION; CONFLICT; CROSS PRESSURE; FAMILY, article on COMPARATIVE STRUCTURE; FASHION; GROUPS, article on THE STUDY OF GROUPS; INTEGRATION, article on CULTURAL INTEGRATION;



INTERACTION, articles on SOCIAL INTERACTION and SOCIAL EXCHANGE; MASS SOCIETY; PRIVACY; PROSTITUTION; TEACHING; and the biographies of LUKÁCS; SMALL; WIESE.]

## WORKS BY SIMMEL

- (1890) 1910 *Über soziale Differenzierung: Soziologische und psychologische Untersuchungen*. 3d ed. Leipzig: Duncker & Humblot.
- (1892) 1923 *Die Probleme der Geschichtsphilosophie: Eine erkenntnis-theoretische Studie*. 5th ed. Munich and Leipzig: Duncker & Humblot.
- 1896 Superiority and Subordination as Subject-matter of Sociology. *American Journal of Sociology* 2:167-189, 392-415. → Translation of an earlier and much shorter draft of "Über- und Unterordnung," Chapter 3 of Simmel 1908b. An abridged version of the English translation appeared as "Superiority and Subordination in Social Relationships" in Borgatta and Meyer 1956.
- 1897-1898 The Persistence of Social Groups. *American Journal of Sociology* 3:662-698, 829-836; 4:35-60. → A translation of an earlier and much shorter draft of "Die Selbsterhaltung der sozialen Gruppe," Chapter 8 of Simmel 1908b. An abridged version of the English translation appeared in Borgatta and Meyer 1956.
- (1900a) 1958 *Philosophie des Geldes*. 6th ed. Leipzig: Duncker & Humblot.
- (1900b) 1924 Money and Freedom. Pages 552-553 in Robert E. Park and Ernest W. Burgess (editors), *Introduction to the Science of Sociology*. 2d ed. Univ. of Chicago Press. → A translation of pages 351-352 of Simmel 1900a.
- 1902 The Number of Members as Determining the Sociological Form of the Group. *American Journal of Sociology* 8:1-46, 158-196. → Translation of an earlier and shorter draft of "Die quantitative Bestimmtheit der Gruppe," Chapter 2 of Simmel 1908b. An abridged version of the English translation appeared in Borgatta and Meyer 1956.
- 1904 The Sociology of Conflict. *American Journal of Sociology* 9 490-525, 672-689, 798-811. → A translation of an earlier and much shorter draft of Chapter 4 of Simmel 1908b, entitled "Der Streit."
- 1906 The Sociology of Secrecy and of Secret Societies. *American Journal of Sociology* 11:441-498. → Translation of an earlier and shorter draft of "Das Geheimnis und die geheime Gesellschaft," Chapter 5 of Simmel 1908b. An abridged version of the English translation appeared as "Knowledge About Others and Social Relationships" in Borgatta and Meyer 1956.
- (1908a) 1955 *Conflict and The Web of Group Affiliations*. Glencoe, Ill.: Free Press. → These essays appeared originally as "Der Streit" and "Die Kreuzung sozialer Kreise" in Simmel 1908b. "Conflict" was translated by Kurt H. Wolff; "The Web of Group Affiliations," by Reinhard Bendix.
- (1908b) 1958 *Soziologie: Untersuchungen über die Formen der Vergesellschaftung*. 4th ed. Berlin: Duncker & Humblot.
- (1908c) 1924 The Sociological Significance of the "Stranger." Pages 322-327 in Robert E. Park and Ernest W. Burgess (editors), *Introduction to the Science of Sociology*. 2d ed. Univ. of Chicago Press. → A translation of pages 685-691 of Simmel 1908b.
- (1908d) 1924 Sociology of the Senses: Visual Interaction. Pages 356-361 in Robert E. Park and Ernest

W. Burgess (editors), *Introduction to the Science of Sociology*. 2d ed. Univ. of Chicago Press. → A translation of pages 646-651 of Simmel 1908b.

- (1910) 1964 *Hauptprobleme der Philosophie*. 8th ed. Berlin: Gruyter.
- (1917) 1920 *Grundfragen der Soziologie (Individuum und Gesellschaft)*. 2d ed. Berlin and Leipzig: Vereinigung Wissenschaftlicher Verleger; Gruyter.
- (1918a) 1922 *Lebensanschauung: Vier metaphysische Kapitel*. 2d ed. Munich and Leipzig: Duncker & Humblot.
- (1918b) 1926 *Der Konflikt der modernen Kultur: Ein Vortrag*. 3d ed. Munich and Leipzig: Duncker & Humblot.
- Philosophische Kultur: Gesammelte Essays*. 3d ed. Potsdam: Kiepenheuer, 1923. → Contains works first published between 1902 and 1911.
- The Sociology of Georg Simmel*. Edited and translated by Kurt H. Wolff. Glencoe, Ill.: Free Press, 1950. → Contains translations, primarily of excerpts from Simmel 1908b, but also of other items published between 1896 and 1917.

## SUPPLEMENTARY BIBLIOGRAPHY

- BORGATTA, EDGAR F.; and MEYER, HENRY J. (editors) 1956 *Sociological Theory: Present Day Sociology From the Past*. New York: Knopf. → See especially pages 126-158, 180-226, 364-398 for extracts from translations of Simmel's writings that first appeared in the *American Journal of Sociology*.
- COSER, LEWIS A. 1956 *The Functions of Social Conflict*. Glencoe, Ill.: Free Press.
- GASSEN, KURT; and LANDMANN, MICHAEL (editors) 1958 *Buch des Dankes an Georg Simmel: Briefe, Erinnerungen, Bibliographie*. Berlin: Duncker & Humblot.
- LUKÁCS, GYÖRGY (1953) 1954 *Die Zerstörung der Vernunft*. Berlin: Aufbau. → First published in Hungarian.
- PARK, ROBERT E.; and BURGESS, ERNEST W. (editors) (1921) 1924 *Introduction to the Science of Sociology*. 2d ed. Univ. of Chicago Press. → Includes a number of translations of Simmel's writings.
- SPYKMAN, NICHOLAS J. (1925) 1964 *The Social Theory of Georg Simmel*. New York: Russell.
- WEINGARTNER, RUDOLPH H. 1962 *Experience and Culture: The Philosophy of Georg Simmel*. Middletown, Conn.: Wesleyan Univ. Press.
- WOLFF, KURT H. (editor) 1959 *Georg Simmel, 1858-1918: A Collection of Essays, With Translations and a Bibliography*. Columbus: Ohio State Univ. Press.

## SIMONDE DE SISMONDI, J. C. L.

Jean Charles Léonard Simonde de Sismondi (1773-1842), historian and economist, was born in Geneva, Switzerland. His father was a Protestant clergyman who had married the daughter of a well-to-do Genevan family. At the age of 16, when Sismondi was a clerk in Eynard's Bank at Lyons, he witnessed the first tremors of the French Revolution. The revolution had repercussions in Geneva, and the family sought refuge in England, returning in 1794 but moving again soon after-

ward to a farm at Pescia, Italy. After his return to Geneva in 1800, Sismondi wrote his *Tableau de l'agriculture toscane*, which appeared the following year. In 1803 he published his book *De la richesse commerciale*, as a result of which he was offered the chair of political economy at the University of Vilna, with a salary of 6,000 francs. Sismondi declined the offer, having decided to devote his energies to his great *Histoire des républiques italiennes du moyen âge*, the 16 volumes of which appeared over the period 1809–1818.

In 1804/1805, Sismondi accompanied Madame de Staël on her voyage to Italy, along with August Wilhelm von Schlegel and other celebrities who frequented her chateau at Coppet, eight miles northeast of Geneva. In 1807/1808, the same group traveled in Germany and Austria, where the baroness took pleasure in introducing Sismondi to her aristocratic friends. Before long, he began to receive handsome royalties for his history of the Italian republics as well as fees for his numerous articles. The Geneva aristocracy, which had earlier turned its back on Sismondi because, during the Revolution, his father had been seen taking the milk of his cows to market (*Fragments . . .*, p. 70), now offered him a professorship at the university. Again, Sismondi declined.

Unfortunately, a valuable source of information on Sismondi's life and opinions seems to have been lost. It is believed that after his death some of his correspondence was burned by his widow and by the marquess of Bossi, his executor.

Sismondi's reputation as a historian and political writer spread rapidly, and during the Hundred Days he had an interview with Napoleon himself, whose constitutional program he defended. After the Restoration, Sismondi resumed work on his history of the Italian republics and, having seen the last volume through the press, threw himself at once into a still vaster project, his *Histoire des Français*; by the time of his death in 1842, 29 volumes had appeared, and two more were published posthumously.

For the modern student, Sismondi's most important work is his *Nouveaux principes d'économie politique* (1819). Jettisoning the viewpoint he had adopted earlier in his *Richesse commerciale* (1803), Sismondi pointed out in his new book that the supporters of the free-trade doctrine were in pursuit of a false prosperity: the *laissez-faire* policy recommended by Say, Ricardo, Malthus, and McCulloch, and followed by the British government, while increasing material wealth, diminished the amount of enjoyment at the disposal of each individual, and as a result of this policy the

wealthy grew wealthier and the poor grew poorer, more dependent, even destitute. The book was translated at once into Italian, but reaction elsewhere in Europe was slow, except in England, where it took a rather unexpected form.

There is ample evidence that at the time that Sismondi released his *Nouveaux principes*, in the spring of 1819, Malthus had been working on his *Principles of Political Economy* for years, without being sure whether, or when, he was going to be able to finish it. The volume was to have been ready in December 1817 but did not come off the press until April 1820. What speeded up Malthus' work between these two dates was surely the publication of Sismondi's book. A comparison of the two books shows that all of Sismondi's important ideas reappear, sometimes in a slightly different garb, in Malthus' *Principles*, which contains references both to Sismondi and to John Barton, a young British economist who, working independently, had reached similar conclusions [see the biography of BARTON]. Ricardo read Malthus' work as soon as it was available. He noticed the name of Barton, remembered the correspondence he had previously exchanged with him, and revised his own doctrine in the light of the arguments originally presented by Sismondi and Barton and now by Malthus. This resulted in the addition to Ricardo's *Principles of Political Economy and Taxation* (see 1817) of the famous 31st chapter "On Machinery" (added in the 1821 edition), in which the author virtually abandoned the free-trade doctrine he had advocated until then. However, Ricardo was not aware of the full significance of his change of mind, for after a meeting he had with Sismondi at Coppet in 1822 he wrote that "after a long battle we each of us, I believe, remained in the same opinion that we commenced the discussion in" (*Works*, vol. 9, p. 248).

Sismondi was respected by men like Ricardo and Barton; but according to Cavour, who made his acquaintance while in Geneva, Sismondi was not a strong debater and could not put up an effective defense against the attacks of men like the economist Antoine E. Cherbuliez. In later years, bourgeois writers usually commented favorably on Sismondi's literary talent, while glossing over his criticism of free-trade doctrine. Socialist writers made use of this criticism, sometimes without giving credit where it was due. Karl Marx, while borrowing copiously from Sismondi's critical analysis of free trade, attacked him as an "economic romanticist." Lenin also attacked Sismondi, and more violently than Marx did.

From the perspective of the twentieth century



Sismondi has proved to have been more perceptive than most free-traders and more realistic than most nineteenth-century socialists.

G. SOTIROFF

[For the historical context of Sismondi's work, see LAISSEZ-FAIRE; and the biographies of MALTHUS; RICARDO; SAY; for discussion of the subsequent development of Sismondi's ideas, see the biography of MARX.]

#### WORKS BY SIMONDE DE SISMONDI

- 1801 *Tableau de l'agriculture toscane*. Geneva: Paschoud.  
 1803 *De la richesse commerciale: Ou, principes d'économie politique, appliqués à la législation du commerce*. 2 vols. Geneva: Paschoud.  
 (1809-1818) 1815-1818 *Histoire des républiques italiennes du moyen âge*. 2d ed. 16 vols. Paris: Treuttel.  
 (1819) 1951-1953 *Nouveaux principes d'économie politique: Ou, de la richesse dans ses rapports avec la population*. 3d ed. 2 vols. Geneva: Jeheber.  
 1821-1844 *Histoire des français*. 31 vols. Paris.  
*Fragments de son journal et correspondance*. Geneva: Cherbuliez, 1857.  
*Political Economy, and the Philosophy of Government: A Series of Essays*. Selected and translated by M. Mignet. London: Chapman, 1847. → Contains an account of Sismondi's life and writings.

#### SUPPLEMENTARY BIBLIOGRAPHY

- AMONN, ALFRED 1945-1949 *Simonde de Sismondi als Nationalökonom: Darstellung seiner Lehren mit einer Einführung und Erläuterungen*. 2 vols. Bern: Francke.  
 GROSSMANN, HENRYK 1924 *Simonde de Sismondi et ses théories économiques (Une nouvelle interprétation de sa pensée)*. Warsaw: Universitas Liberae Poloniae.  
 LENIN, VLADIMIR I. (1896) 1951 *A Characterization of Economic Romanticism: Sismondi and Our Native Sismondists*. Moscow: Foreign Language Publishing House. → First published as "Sotsial'no-ekonomicheskiiya vozreniia Simonda de-Sismondi" in *Russkoe bogatstvo*.  
 RICARDO, DAVID (1817) 1962 *Principles of Political Economy and Taxation*. London: Dent; New York: Dutton. → A paperback edition was published in 1963 by Irwin.  
 RICARDO, DAVID *Works*. Volumes 6-9. Cambridge Univ. Press, 1951-1952.  
 SALIS, JEAN R. DE 1932 *Sismondi, 1773-1842: La vie et l'oeuvre d'un cosmopolite philosophe*. Paris: Champion.  
 SOTIROFF, GEORGES 1945 *Ricardo und Sismondi: Eine aktuelle Auseinandersetzung über Nachkriegswirtschaft vor 120 Jahren*. Zurich and New York: Europa.

#### SIMONS, HENRY C.

Henry Calvert Simons (1899-1946), American economist, spent most of his professional career at the University of Chicago, teaching in the department of economics and then becoming the first professor of economics in the law school. He de-

veloped a unified and comprehensive set of principles of economic policy for the United States and applied these principles to the changing problems of his times.

Simons called his philosophy libertarian; the essential requirement of a good political and economic order is that it protect, promote, and respond to the free choices of individuals. However, what is distinctive about his philosophy is the combination of this idea with two others: First, a merely negative, "hands-off" policy of government is not sufficient to maintain the free economic order and realize its benefits; the government must undertake certain limited but important affirmative actions. (The title of his first work to gain national attention, "A Positive Program for Laissez-faire," 1934, already conveyed this idea, and it recurs throughout his writing.) Second, equality is a value only a little below freedom.

Simons' important writings began to appear in 1933 and 1934, near the bottom of the great depression. The failure of "free enterprise" and laissez-faire had greatly strengthened support for government intervention in economic activity. There were no apparent limits to the character and scope of these interventions, less because of any widespread acceptance of a policy of extensive government control than because of the absence of any principles for distinguishing between controls that were necessary, appropriate, or feasible and those that were not.

Simons proposed a set of principles for making this distinction: It is the responsibility of the government to maintain a framework within which a free economy can successfully operate and to do so by means that are general in their application, impersonal, and based on law rather than administrative discretion. But the government should not intervene in the heart of the economic process—in the determination of relative prices and wages and in production and investment decisions. Specifically, Simons emphasized the government's responsibility for maintaining appropriate monetary and competitive conditions.

Simons believed that the failure of the government to discharge its responsibility for maintaining reasonable stability in the supply of money had been one of the main causes of the severity of the depression. He was an early leader, among American economists, in advocating monetary expansion, to be achieved by means of budget deficits, as the road to recovery from the depression. But he did not believe that it was sufficient merely to assign to the government the responsibility of using its monetary, fiscal, and debt-management policies

to achieve some vaguely defined goal of economic stability. Reform of the banking system was also necessary, and he urged that the money supply be based on a 100 per cent reserve of government obligations. Also, he argued the case for a statutory rule of monetary policy that would govern the conduct of the monetary authorities. Which of several reasonable rules was chosen was less important than the existence of some objective standard known to the public and binding on the authorities.

Simons often argued against what he considered excessively loose, discretionary, and expansionist prescriptions by overenthusiastic advocates of positive fiscal and monetary policy. Yet, the more important fact is that Simons, the champion of free enterprise and nonintervention, did accept and emphasize the responsibility of the government to use its monetary and fiscal policies affirmatively to stabilize the economy. This allowed Simons and his followers to influence policy more than they could have done if they had denied government responsibility. Simons was among the first to stress that if federal tax rates were unchanged, the resulting decline of revenues in recessions and rise of revenues in booms would have a strong stabilizing influence on the economy; that varying the revenue side of the budget might be a more effective fiscal instrument than varying the expenditure side; and that the way in which the government financed its debt would be a significant independent economic factor. These ideas all became important in postwar fiscal thinking and fiscal policy. Thus, Simons' writings made a major contribution to the emergence of a postwar consensus on fiscal and monetary policy in the United States, not only among economists but also among persons in government and in the private economy whose opinions were influential.

While Simons believed that monetary stability is a necessary background condition for the effective operation of the free economy, he also felt that competition is the active driving force of the economy. He relied upon competition to organize production efficiently; to promote technological advance and economic growth; to permit high employment without inflation; to provide individuals, as consumers and as sellers of productive services, with the real alternatives that are the essence of freedom; to yield "commutative justice," in the sense of economic rewards proportional to productive contributions; to prevent extreme inequality of power or income; and to minimize the need for government intervention to settle conflicts between private interests.

In his early writings Simons seemed to believe

that the degree of competition in the United States was far from adequate to serve these purposes, chiefly because corporations had been permitted to grow to excessive size. Later, Simons' view of the problem changed. He came to believe that it was the trend toward monopoly, rather than the degree of monopoly already achieved, that was the great danger. Moreover, the natural force of competition was strong, so strong that private arrangements, unaided by government, could only temporarily and partially suppress it. Therefore, the first requirement was to prevent government from supporting monopoly.

In his last writings, including "Some Reflections on Syndicalism" (1944), which became the classic statement of the case against trade unions, Simons argued that unions were the principal threat to the competitive, and therefore free, society. Simons offered no confident remedy, only the hope that awareness of the danger would make it possible to limit the growth of union power before it was too late.

The part of his program that Simons elaborated in greatest detail was the tax system. His two longest (although relatively short) works, *Personal Income Taxation* (1938) and *Federal Tax Reform* (written in 1943, published 1950), were on this subject. Simons' writing on taxation largely spelled out the implications of a few basic ideas, the one most closely identified with his name being the definition of income, for tax purposes, as the sum of consumption plus the change in the value of net assets owned. Other ideas fundamental to his thinking about taxes were (a) that all taxation is taxation of individuals and must be judged by its effects on individuals; (b) that the allocation of income to short periods of time, like a year, is arbitrary and causes injustice and inefficiency; and (c) that in taxation, as in other policies, the government should avoid accidental or deliberate interference with the system of relative prices and wages. Rigorous application of these ideas enabled Simons to cut through a great deal of sophisticated argument about such matters as capital gains, gifts and bequests, corporate profits, and tax-exempt securities. Twenty years after his death, the considerations contained in Simons' work were still basic to the continuing effort to rationalize the federal income tax system.

It was in his tax proposals that Simons' stress on equality as a social goal found its most immediate expression. Simons did not believe that policy for reducing inequality should be confined to creation of a progressive tax system, but he did believe that correction of the tax system had to



come first. As he said, "There are remarkable opportunities for extending the range of socialized consumption (medical services, recreation, education, music, drama, etc.) and especially for extending the range of social welfare activities. The prospects in these directions, however, must remain somewhat unattractive as long as the expenditures involved must be covered by the kinds of taxes on which we have relied in the past—that is, so long as what the government gives to the masses with one hand is largely taken away with the other" ([1934] 1948, pp. 68–69).

HERBERT STEIN

[See also FISCAL POLICY; MONETARY POLICY; TAXATION, article on PERSONAL INCOME TAXES; and the biography of KNIGHT.]

#### WORKS BY SIMONS

- (1934) 1948 *A Positive Program for Laissez-faire: Some Proposals for a Liberal Economic Policy*. Pages 40–77 in Henry C. Simons, *Economic Policy for a Free Society*. Univ. of Chicago Press.
- 1938 *Personal Income Taxation: The Definition of Income as a Problem of Fiscal Policy*. Univ. of Chicago Press.
- (1943) 1950 *Federal Tax Reform*. Univ. of Chicago Press. → Written in 1943; published posthumously in 1950.
- 1944 *Some Reflections on Syndicalism*. *Journal of Political Economy* 52:1–25.
- Economic Policy for a Free Society*. Univ. of Chicago Press, 1948. → Contains writings first published between 1934 and 1945 and a complete bibliography.

#### SUPPLEMENTARY BIBLIOGRAPHY

- DAVENPORT, JOHN 1946 *The Testament of Henry Simons*. *University of Chicago Law Review* 14:5–14. → First published in *Fortune*, in September 1946.
- DIRECTOR, AARON 1946 *Simons on Taxation*. *University of Chicago Law Review* 14:15–20.
- KATZ, WILBER G. 1946 *Economics and the Study of Law: The Contribution of Henry C. Simons*. *University of Chicago Law Review* 14:1–4.
- LEWIS, H. GREGG 1946 [Obituary of] Henry C. Simons. *American Economic Review* 36:668–669.

## SIMULATION

- I. INDIVIDUAL BEHAVIOR
- II. ECONOMIC PROCESSES
- III. POLITICAL PROCESSES

Allen Newell and  
Herbert A. Simon  
Irma Adelman  
Charles F. Hermann

### I

#### INDIVIDUAL BEHAVIOR

"Simulation" is a term now generally employed to denote an approach to the construction of theories that makes essential use of computers and computer-programming languages. In some appli-

cations simulation techniques are used in investigating formal mathematical theories—for example, stochastic learning theories. In other applications, the theoretical models of the systems that are simulated are essentially nonquantitative and non-numerical.

We may define simulation more specifically as a method for analyzing the behavior of a system by computing its time path for given initial conditions, and given parameter values. The computation is said to "simulate" the system, because it moves forward in time step by step with the movement of the system it is describing. For the purposes of simulation, the system must be specified by laws that define its behavior, during any time interval, in terms of the state it was in at the beginning of that interval.

Simulation has always been an important part of the armory of applied mathematics. The introduction of modern computers, however, has so reduced the cost of obtaining numerical solutions to systems of equations that simulation has taken on vastly increased importance as an analytic tool. As a consequence, the relative balance of advantage has shifted from simplifying a theory in order to make it solvable toward retaining complexity in order to increase its accuracy and realism. With simulation and other techniques of numerical analysis, computers permit the study of far more complicated systems than could have been investigated at an earlier time.

Simulation with mathematical theories. Consider first the simulation of systems that are described by mathematical models (Newell & Simon 1963a, pp. 368–373). A model consists of a system of equations containing variables, literal constants (often called parameters), and numerical constants. The system of equations may be solved symbolically, by expressing the several variables explicitly in terms of the literal and numerical constants; or the system may be solved numerically, by substituting particular numerical values for the literal constants and then solving the equations for these special cases. For many complex systems of equations, symbolic solutions cannot be obtained, only numerical solutions. The former are clearly preferred when obtainable, for once a symbolic solution has been found, the special cases are readily evaluated by substituting numerical values for the parameters in the solutions instead of in the original equations.

To take a highly simplified example, a learning theory might postulate that the rate at which learning occurs is proportional to the amount of material

remaining to be learned. If, for simplicity of exposition, we divide time into discrete intervals and define the variable  $x(t)$  as the amount of material that has been learned up to the end of the  $t$ th time interval, then the theory can be expressed by a simple difference equation:

$$(1) \quad x(t+1) - x(t) = a[b - x(t)],$$

where  $a$  is a literal constant that we can name intelligence for this learning task, and  $b$  is a literal constant that represents the total amount to be learned. The symbolic solution to this simple system is well known to be:

$$(2) \quad x(t) = b - [b - \bar{x}(0)](1 - a)^t,$$

where  $\bar{x}(0)$  is a literal constant denoting the amount that had already been learned up to the time  $t = 0$ . Particular numerical solutions can now be found by assigning numerical values  $a$ ,  $b$ , and  $\bar{x}(0)$  and by solving (2) for  $x(t)$  for any desired values of  $t$ . Alternatively, numerical solutions—learning curves—for the system can as readily be computed for given  $a$ ,  $b$ , and  $\bar{x}(0)$  directly from equation (1), transferring the  $x(t)$  term to the right-hand side and solving successively for  $x(1)$ ,  $x(2)$ ,  $x(3)$ ,  $\dots$ . This latter procedure—solving (1) numerically—would be called simulation.

The advantages of simulation as a technique for analyzing dynamic systems are not evident from the excessively simple equation (1). Suppose, however, that  $a$ , instead of being a literal constant, designated some function,  $a[b, x(t)]$ , of  $b$  and  $x(t)$ . Only for very special cases could (1) then be solved symbolically to obtain an explicit expression, like (2), for  $x(t)$ . In general, when we wish to know the time paths of systems expressed by such equations, numerical solution for special cases—that is, simulation—is our main recourse.

A great many of the mathematical theories currently used for the study of learning and other individual behavior are probabilistic or stochastic in character. Such theories pose formidable mathematical problems, and explicit symbolic solutions for their equations are seldom obtainable. Unlike a deterministic system, the solution of the equations of a stochastic system does not specify a single time path for the system but assigns a probability distribution to all possible paths of the system. In some cases, certain properties of this probability distribution—certain means and variances—can be obtained symbolically, but the specification of the entire probability distribution can seldom be obtained symbolically. In order to estimate the parameters of the distribution numerically, the

system can be simulated by Monte Carlo techniques. That is, random variables can be introduced into the model, and the equations analogous to (1) solved a number of times, each time with different values of the random variables assigned and using the appropriate probabilities. The numerical solutions to (1) will trace out *possible* paths of the system, and the probability that a particular path will be followed in the simulation will be proportional to its probability under the assumptions of the model. Then, from the numerical results of the Monte Carlo simulation, numerical estimates can be made of the parameters of the probability distribution. [See RANDOM NUMBERS.]

**Information-processing theories.** The greatest use of simulation for the study of individual behavior has involved so-called information-processing theories, which are different in important respects from classical mathematical theories. Simulation has exactly the same relation to information-processing theories as to other formal theories—it is a method for discovering how the system described by the theory will behave, in particular circumstances, over a period of time. However, almost no methods other than simulation exist for investigating information-processing theories. For this reason, a phrase like “simulation of cognitive processes” often refers both to an information-processing theory of cognitive processes and to the investigation of the theory by simulation, in the narrower sense of the term. [See INFORMATION THEORY.]

Information-processing theories of individual behavior take the form of computer programs, usually written in programming languages called list-processing languages, especially devised for this purpose. The information-processing theories undertake to explain how a system like the human central nervous system carries out such complex cognitive processes as solving problems, forming concepts, or memorizing. The explanation is sufficiently detailed to predict—by computer simulation—behavior in specific problem situations. That is to say, the theory aims at predicting, not merely some quantitative aspects of behavior (number of errors in a learning situation, for example), but the actual concrete behaviors and verbal outputs of subjects placed in the very same situation and confronted with the identical task. [See CONCEPT FORMATION; PROBLEM SOLVING.]

A specific example will help make clear exactly what is meant by this. Clarkson (1963) has constructed an information-processing theory to ex-



plain how a bank officer selects portfolios of securities for trust funds. The trust officer, who has been told the purpose of the trust and the amount of money available for investment, prepares a list of stocks and bonds that can be bought with this sum. If the computer program is provided (as input) the same information about the purpose of the trust and its size, it will also produce (as output) a list of stocks and bonds. A first test of the program, as a theory, is whether it will pick the same list of companies and the same number of shares in each as the trust officer. In an actual test, for example, the program predicted the trust officer would buy 60 shares of General American Transportation, 50 Dow Chemical, 10 IBM, 60 Merck and Company, and 45 Owens-Corning Fiberglas. The trust officer in fact bought 30 Corning Glass, 50 Dow Chemical, 10 IBM, 60 Merck and Company, and 50 Owens-Corning Fiberglas.

A second and more severe test of the theory is whether, in reaching its final choices, it goes through the same stages of analysis, weighs the same factors, considers the same alternatives, as the human it is simulating. As one technique for making this kind of test, the human subject is asked to perform the task while thinking aloud, and his stream of verbalizations (protocol) is recorded. The human protocols are then compared with the *trace* produced by the computer program while it is performing the same task (for an example of this technique, which Clarkson also used, see Newell & Simon 1963b).

Omar K. Moore and Scarvia B. Anderson (1954) devised a task that required the subject to "recode" certain symbolic expressions. One problem was to recode  $L1: (r \supset -p) \cdot (-r \supset q)$  into  $L0: -(-q \cdot p)$ . (For purposes of the experiment and of this example, these strings of symbols can be treated as "code messages," whose meaning need not be known. The expression  $L1$  may be read "parenthesis  $r$  horseshoe minus  $p$  parenthesis dot parenthesis minus  $r$  horseshoe  $q$  parenthesis.") The recoding had to be carried out according to certain rules, which were numbered from 1 to 12. At one point in his thinking-aloud protocol, a subject said, "Now I'm looking for a way to get rid of the horseshoe inside the two brackets that appear on the left and right side of the equation. And I don't see it. Yeah, if you apply rule 6 to both sides of the equation, from there I'm going to see if I can apply rule 7."

The same recoding task, recoding  $L1$  into  $L0$  according to the rules specified by Moore and Anderson, was given to a program, the General Problem Solver (GPS), expressing an information-

processing theory of human problem solving. One portion of the trace produced by the program read as follows:

Goal: Apply rule 7 to  $L1$ .

Subgoal: Change "horseshoe" to "wedge" in left side of  $L1$ .

Subgoal: Apply rule 6 to the left side of  $L1$ .

Result:  $L4: (\neg r \vee \neg p) \cdot (\neg r \supset q)$ .

Subgoal: Apply rule 7 to  $L4$ .

Subgoal: Change "horseshoe" to "wedge" in right of  $L4$ .

Subgoal: Apply rule 6 to right side of  $L4$ .

Part of the test of the adequacy of GPS as a theory of human problem solving would be to decide how closely the segment of the human protocol reproduced above corresponded to the segment of the computer trace. For example, in the illustrative fragment cited here, both the human subject and GPS went down a blind alley, for rule 7 turned out to be not applicable to the expression recoded by rule 6. A theory of human problem solving must predict and explain the mistakes people make, as well as their ability sometimes to achieve solutions.

Information-processing theories can also be tested in more orthodox ways than by comparing protocols with traces. A program can be simulated in an experimental design identical with one that has been employed for human subjects. Statistics from the computer runs can be compared with the statistics of the human performance. This is the principal means that has been used to test the Elementary Perceiver and Memorizer (EPAM), an information-processing theory of human rote learning that will be described briefly later (Feigenbaum 1963). By having EPAM learn nonsense syllables by the standard serial anticipation method, at various simulated memory drum speeds and for lists of different lengths, quantitative predictions were made of the shape of the so-called serial position curve (relative numbers of errors during the learning process for different parts of the list of syllables). These predictions showed excellent agreement with the published data from human experiments. In other experiments in the learning of paired nonsense syllables, EPAM predicted quantitatively the effects upon learning rate of such variables as degree of familiarity with the syllables and degree of similarity between syllables. These predictions also corresponded closely with the published data from several experiments. [See FORGETTING and LEARNING, article on VERBAL LEARNING.]

In summary, an information-processing theory is expressed as a computer program which, exactly like a system of difference or differential equations, predicts the time path of a system from given initial conditions for particular values of the system parameters. The theory predicts not merely gross quantitative features of behavior but the actual stream of symbolic outputs from the subject. Such theories can be subjected to test in numerous ways; among others, by comparing the behaviors, including the thinking-aloud protocols, of subjects with the computer traces produced by the simulating programs.

**Evaluation.** The examples reveal both some of the strengths and some of the difficulties in the information-processing approach to theory construction and theory testing. On the positive side, the theories are written in languages that are capable of representing stimuli and responses directly and in detail, without an intermediate stage of translation into mathematical form. This has the further advantage of avoiding virtually all problems of stimulus scaling. The stimuli themselves, and not scales representing some of their gross characteristics, serve as inputs to the theory. Although the computer traces produced by the simulations are not in idiomatic English, the comparison of trace with protocol can be handled by relatively unambiguous coding techniques.

One of the main difficulties of the approach is closely related to one of its strengths—the detail of its predictions. Since different human subjects do not behave in exactly the same way in identical task situations, a simulation that predicts correctly the detail of behavior of one individual will predict the detail of behavior of others incorrectly. Presumably, a theory of individual behavior should consist of general statements about how human beings behave, rather than statements about how a particular human being behaves. At a minimum we would want to require of a theory that modifications to fit the behaviors of different subjects should change only relatively superficial features of the theory—values of particular parameters, say—and that they not necessitate fundamental reconstruction of the whole simulation program. It remains to be seen how fully this requirement will be met by information-processing theories. Up to the present, only a modest amount of investigation has been made of the possibilities of creating variants of simulation programs to fit different subjects.

Generality is needed, too, in another direction. Most of the early simulation programs were constructed to explain human performance in a single

task—discovering proofs for theorems in symbolic logic, say, or making moves in chess. It seems unreasonable to suppose that the programs a human uses to solve problems in one task environment are entirely specific to that environment and separate and independent from those he uses for different tasks. The fact that people can be scaled, even roughly, by general intelligence argues against such total specialization of skills and abilities. Hence, an important direction of research is to construct theories that can be applied over a wide range of tasks.

An example of such a theory is GPS, already mentioned. GPS can simulate behavior in any problem environment where the task can be symbolized as getting from a given state of affairs to a desired state of affairs. A wide range of problem-solving tasks can be expressed in this format. (To say that GPS can simulate behavior in any such environments means, not that the program will predict human behavior correctly and in detail, but that the program can at least be made to operate and produce a prediction, a trace, in that environment, for comparison with the human behavior.)

Thus, GPS has actually been given tasks from about a dozen task environments—for example, discovering proofs for theorems in logic, solving the missionaries-and-cannibals puzzle, and solving trigonometric identities—and there is every reason to suppose it can handle a wide range of others.

**Levels of explanation.** There has been great diversity of opinion in individual psychology as to the appropriate relation between psychological theory and neurophysiology. One extreme of the behaviorist position holds that the laws of psychology should state functional relations between the characteristics of stimulus situations, as the independent variables, and response situations, as the dependent variables, with no reference to intervening variables. A quite different position holds that explanation in psychology should relate behavior to biological mechanisms and that the laws of psychology are essentially reducible to laws of neurophysiology.

Information-processing theories represent a position distinct from either of these (Simon & Newell 1964). Unlike the extreme behaviorist theories, they make strong assumptions about the processes in the central nervous system (CNS) that intervene between stimulus and response and endeavor to explain how the stimulus produces the response. On the other hand, the information-processing theories do not describe these intervening processes in terms of neurophysiological mechanisms but postulate much grosser elements: symbol struc-



tures and elementary information processes that create and modify such structures and are, in turn, controlled by the structures.

Behaviorist theories may be called one-level theories, because they refer only to directly observable phenomena. Neurophysiological theories are two-level theories, for they refer both to behavioral events and underlying chemical and biological mechanisms. Information-processing theories are at least three-level theories, for they postulate elementary information processes to explain behavior, with the suggestion that the elementary information processes can subsequently be explained (in one or more stages of reduction) in chemical and biological terms. [See LEARNING, *article on NEUROPHYSIOLOGICAL ASPECTS*; and NERVOUS SYSTEM.]

Elementary information processes and symbol structures play the same role, at this intermediate level of explanation, as that played by chemical reactions, atoms, and molecules in nineteenth-century chemistry. In neither case is it claimed that the phenomena cannot be reduced, at least in principle, to a more microscopic level (neurophysiology in the one case, atomic physics in the other); it is simply more convenient to have a division of labor between psychologists and neurophysiologists (as between chemists and atomic physicists) and to use aggregative theories at the information-processing level as the bridge between them.

**Some specific theories.** The remainder of this article will be devoted to an account of the psychological substance of some current information-processing theories of cognition. Because space prohibits a complete survey of such theories, the examples will be restricted to problem solving, serial pattern recognition, and rote memory processes. Theories of other aspects of pattern recognition and concept formation will be omitted (Hunt 1962; Uhr 1966), as will be theories proposing a parallel, rather than serial, organization of cognitive processes (Reitman 1965).

The theories to be described are incorporated in a number of separate programs, which have not yet been combined into a single information-processing theory of "the whole cognitive man." Nevertheless, the programs have many similar components, and all incorporate the same basic assumptions about the organization and functioning of the CNS. Hence, the theories are complementary and will be discussed here, not as separate entities, but as components of a theory of cognition (Simon & Newell 1964).

The basic assumptions about the organization and functioning of the CNS are these:

(1) The CNS contains a *memory*, which stores

symbols (discriminable patterns) and composite structures made up of symbols. These composite structures include *lists* (ordered sets of symbols, e.g., the English alphabet) and *descriptions* (associations between triads of symbols). The relation "black, opposite, white," to be translated "white is the opposite of black," is an example of a description.

(2) The CNS performs certain *elementary processes* on symbols: storing symbols, copying symbols, associating symbols in lists and descriptions, finding symbols on lists and in descriptions, comparing symbols to determine whether they are identical or different.

(3) The elementary processes are organized hierarchically into *programs*, and the CNS incorporates an *interpretive process* capable of executing such programs—determining at each stage what elementary process will be carried out next. The interpretation is serial (not all information-processing theories share this assumption, however), one process being executed at a time.

The memory, symbol structures, elementary processes, and interpreters are the information-processing mechanisms in terms of which observed human behavior is to be explained. Programs constructed from these mechanisms are used to simulate problem solving, memorizing, serial pattern recognizing, and other performances.

**Problem-solving processes.** A central task of any theory of human problem solving is to explain how people can find their way to solutions while exploring enormous "mazes" of possible alternative paths. It has been calculated that if a chess player were to consider all the possible outcomes of all possible moves, he would have to examine some  $10^{120}$  paths. In fact he does not do this—obviously he could not, in any event—but conducts a highly selective search among a very much smaller number of possibilities. Available empirical data indicate that a problem solver usually explores well under a hundred paths in this and other problem-solving "mazes."

Experiments based on information-processing theories have shown that such selective searches can be accomplished successfully by programs incorporating rules of thumb, or heuristics, for determining which paths in the maze are likely to lead to solutions (Feigenbaum & Feldman 1963, part 1, secs. 2, 3). Some of the heuristics are very specific to a particular task environment. As a person becomes skilled in such an environment, he learns to discriminate features of the situation that have diagnostic value, and he associates with those features responses that may be appropriate in situa-

tions of the specified kind. Thus, an important component in specific skills is a "table of connections," stored in memory, associating discriminable features with possibly relevant actions. For example, a chess player learns to notice when the enemy's king is exposed, and he learns to consider certain types of moves (e.g., "checking" moves) whenever he notices this condition. [See ATTENTION.]

Much of the behavior of problem solvers in chess-playing and theorem-proving tasks can be explained in terms of simple "feature-noticing" programs and corresponding tables of connections, or associations. But the programs must explain also how these components in memory are organized into effective programs for selective search. These organizing programs are relatively independent of the specific kind of problem to be solved but are applicable instead (in conjunction with the noticing processes and associations) to large classes of task environments. They are general heuristics for problem solving.

The central GPS heuristic is an organization of elementary processes for carrying out *means-end analysis*. These processes operate as follows: the symbolized representations of two situations are compared; one or more differences are noted between them; an operator is selected among those associated with the difference by the table of connections; and the operator is applied to the situation that is to be changed. After the change has been made, a new situation has been created, which can be compared with the goal situation, and the process can then be repeated.

Consider, for example, a subject trying to solve one of the Moore-Anderson "recoding" problems. In this case, the problem is to change  $L1: r \cdot (\neg p \supset q)$  to  $L0: (q \vee p) \cdot r$ . The subject begins by saying, "You've got to change the places of the  $r$  to the other side and you have to change the . . . minus and horseshoe to wedge, and you've got to reverse the places from  $p$  and  $q$  to  $q$  and  $p$ , so let's see. . . . Rule 1 (i.e.,  $A \cdot B \rightarrow B \cdot A$ ) is similar to it because you want to change places with the second part. . . ."

The means-end chain in this example includes the following: notice the difference in order between the terms of  $L1$  and  $L0$ ; find an operator (rule 1) that changes the order.

Another kind of organization, which may be called *planning and abstraction*, is also evident in problem solving. In planning, some of the detail of the original problem is eliminated (abstracted). A solution is then sought for the simplified problem, using means-end analysis or other processes.

If the solution is found for the abstract problem, it provides an outline or plan for the solution of the original problem with the detail reinserted.

The programs for means-end analysis and for planning, each time they are activated, produce a "burst" of problem-solving activity directed at a particular subgoal or organization of subgoals. The problem-solving program also contains processes that control and organize these bursts of activity as parts of the total exploratory effort. [See PROBLEM SOLVING.]

**Rote memory processes.** The noticing processes incorporated in the problem-solving theory assume that humans have subprograms for discriminating among symbol structures (e.g., for noticing differences between structures) and programs for elaborating the fineness of these discriminations. They also assume that associations between symbol structures can be stored (e.g., the table of connections between differences and operators). Discrimination and association processes and the structures they employ constitute an important part of the detail of problem-solving activity but are obscured in complex human performances by the higher-level programs that organize and direct search.

In simpler human tasks—memorizing materials by rote—the organizing strategies are simpler, and the underlying processes account for most of the behavioral phenomena. It has been possible to provide a fairly simple information-processing explanation for the acquisition of discriminations and associations (Feigenbaum 1963). (Our description follows EPAM, the program for rote learning mentioned above.) In the learning situation, there develops in memory (a) a "sorting net," for discriminating among stimuli and among responses, and (b) compound symbol structures, stored in the net, that contain a partial representation, or "image," of the stimulus as one component and a partial image of the response as another.

When stimuli or responses are highly similar to each other, greater elaboration of the sorting net is required to discriminate among them, and more detail must be stored in the images. A stimulus becomes "familiar" as a result of repeated exposure and consequent elaboration of the sorting net and its stored image. We have already noted that these simple mechanisms have been shown to account, quantitatively as well as qualitatively, for some of the main features of learning rates that have been observed in the laboratory. [See LEARNING, article on DISCRIMINATION LEARNING.]

**Recognition of periodic patterns.** As an example of an information-processing explanation of



some of the phenomena of concept formation, we consider how, according to such a theory, human subjects detect and extrapolate periodic patterns—e.g., the letter pattern *A B M C D M . . .* (Simon & Newell 1964, pp. 293–294).

At the core of the explanation is the hypothesis that in performing these tasks humans make use of their ability to detect a small number of basic relations and operations: for example, the relations “same” and “different” between two symbols or symbol structures; the relation “next” on a familiar alphabet; and the operations of finding the difference between and the quotient of a pair of numbers. (These discriminations are not unlike those required as components in problem-solving programs to detect features of symbol structures and differences between them.) With a relatively simple program of these processes, the pattern in a sequence can be detected, the pattern can be described by a symbol structure, and the pattern description can be used by the interpretive process to extrapolate. In the example *A B M C D M*, the relation “next” is detected in the substring *A B C D*, and the relation “same” in the *M*’s. These relations are used to describe a three-letter pattern, whose next member is the “next” letter after *D*, i.e., *E*.

The information-processing theory of serial pattern detection has proved applicable to a wide range of cognitive tasks—including the Thurstone Letter Series Completion Test, symbol analogies tests, and partial reinforcement experiments. [See CONCEPT FORMATION; see also Feldman 1963.]

This article has described the use of simulation in constructing and testing theories of individual behavior, with particular reference to nonnumerical information-processing theories that take the form of computer programs. Examples of some of the current theories of problem solving, memorizing, and pattern recognition were described briefly, and the underlying assumptions about the information-processing characteristics of the central nervous system were outlined. Various techniques were illustrated for subjecting information-processing theories to empirical test.

ALLEN NEWELL AND HERBERT A. SIMON

[Directly related are the entries DECISION MAKING, especially the article on PSYCHOLOGICAL ASPECTS; INFORMATION THEORY; MODELS, MATHEMATICAL; PROBLEM SOLVING. Other relevant material may be found in LEARNING THEORY; MATHEMATICS; NERVOUS SYSTEM; PROGRAMMING; REASONING AND LOGIC; THINKING.]

## BIBLIOGRAPHY

- CLARKSON, GEOFFREY P. E. 1963 A Model of the Trust Investment Process. Pages 347–371 in Edward A. Feigenbaum and Julian Feldman (editors), *Computers and Thought*. New York: McGraw-Hill.
- FEIGENBAUM, EDWARD A. 1963 The Simulation of Verbal Learning Behavior. Pages 297–309 in Edward A. Feigenbaum and Julian Feldman (editors), *Computers and Thought*. New York: McGraw-Hill.
- FEIGENBAUM, EDWARD A.; and FELDMAN, JULIAN (editors) 1963 *Computers and Thought*. New York: McGraw-Hill.
- FELDMAN, JULIAN 1963 Simulation of Behavior in the Binary Choice Experiment. Pages 329–346 in Edward A. Feigenbaum and Julian Feldman (editors), *Computers and Thought*. New York: McGraw-Hill.
- GREEN, BERT F. JR. 1963 *Digital Computers in Research: An Introduction for Behavioral and Social Scientists*. New York: McGraw-Hill.
- HUNT, EARL B. 1962 *Concept Learning: An Information Processing Problem*. New York: Wiley.
- MILLER, GEORGE A.; GALANTER, E.; and PRIBRAM, K. H. 1960 *Plans and the Structure of Behavior*. New York: Holt.
- MOORE, OMAR K.; and ANDERSON, SCARVIA B. 1954 Modern Logic and Tasks for Experiments on Problem Solving Behavior. *Journal of Psychology* 38:151–160.
- NEWELL, ALLEN; and SIMON, HERBERT A. 1963a *Computers in Psychology*. Volume 1, pages 361–428 in R. Duncan Luce, Robert Bush, and Eugene Galanter (editors), *Handbook of Mathematical Psychology*. New York: Wiley.
- NEWELL, ALLEN; and SIMON, HERBERT A. 1963b GPS, a Program That Simulates Human Thought. Pages 279–293 in Edward A. Feigenbaum and Julian Feldman (editors), *Computers and Thought*. New York: McGraw-Hill.
- REITMAN, WALTER R. 1965 *Cognition and Thought: An Information-processing Approach*. New York: Wiley.
- SIMON, HERBERT A.; and NEWELL, ALLEN 1964 Information Processing in Computer and Man. *American Scientist* 52:281–300.
- UHR, LEONARD (editor) 1966 *Pattern Recognition*. New York: Wiley.

## II

### ECONOMIC PROCESSES

Simulation is at once one of the most powerful and one of the most misapplied tools of modern economic analysis. “Simulation” of an economic system means the performance of experiments upon an analogue of the economic system and the drawing of inferences concerning the properties of the economic system from the behavior of its analogue. The analogue is an idealization of a generally more complex real system, the essential properties of which are retained in the analogue.

While this definition is consistent with the denotation of the word, the connotation of simulation among economists active in the field today is much more restricted. The term “simulation” has been generally reserved for processes using a physical

or mathematical analogue and requiring a modern high-speed digital or analogue computer for the execution of the experiments. In addition, most economic simulations have involved some stochastic elements, either as an intrinsic feature of the model or as part of the simulation experiment field. Thus, while a pencil-and-paper calculation on a two-person Walrasian economy would, according to the more general definition, constitute a simulation experiment, common usage of the term would require that, to qualify as a simulation, the size of the model must be large enough, or the relationships complex enough, to necessitate the use of a modern computing machine. Even the inclusion of probabilistic elements in the pencil-and-paper game would not suffice, in and of itself, to transform the calculations into a bona fide common-usage simulation. In order to avoid confusion, the term "simulation" will be used hereafter in its more restrictive sense.

To clarify the concept of simulation, it might be useful to indicate how the simulation process is related, on the one hand, to the analytic solution of mathematical models of an economy, and on the other, to the construction of mathematical and econometric models of economic systems.

The relationship between simulation and econometric and mathematical formulations of economic theories is quite intimate. It is these descriptions of economic processes that constitute the basic inputs into a simulation. After a mathematical or econometric model is translated into language a computing machine can understand, the behavior of the model, as described by the machine output, represents the behavior of the economic system being simulated. The solutions obtained by means of simulation techniques are quite specific. Given a particular set of initial conditions, a particular set of parameters, and the time period over which the model is to be simulated, a single simulation experiment yields a particular numerically specified set of time paths for the endogenous variables (the variables whose values are determined or explained by the model). A variation in one or more of the initial conditions or parameters requires a separate simulation experiment which provides a different set of time paths. Comparisons between the original solution and other solutions obtained under specific variations in assumptions can then be used to infer some of the properties of the relationships between input and output quantities in the system under investigation. In general, only very partial inferences concerning these relationships can be drawn by means of simulation experiments. In

addition, the results obtained can be assumed to be valid only for values of the parameters and initial conditions close to those used in the simulation experiments. By contrast, traditional mathematical approaches for studying the implications of an economic model produce general solutions by deductive methods. These general solutions describe, in functional form, the manner in which the model relates the endogenous variables to any set of initial conditions and parameters and to time.

The models formulated for a simulation experiment must, of course, represent a compromise between "realism" and tractability. Since modern computers enable very large numbers of computations to be performed rapidly, they permit the step-by-step solution of systems that are several orders of magnitude larger and more complicated than those that can be handled by the more conventional techniques. The representation of economic systems to be investigated with the aid of simulation techniques can therefore be much more complex; there are considerably fewer restrictions on the number of equations and on the mathematical forms of the functions that can be utilized. Simulation, therefore, permits greater realism in the extent and nature of the feed-back mechanisms in the stylized representation of the economy contained in the econometric or mathematical model.

The impetus for the use of simulation techniques in economics arises from three major sources. First, both theory and casual observation suggest that an adequate description of the dynamic behavior of an economy must involve complex patterns of time dependencies, nonlinearities, and intricate interrelationships among the many variables governing the evolution of economic activity through time. In addition, a realistic economic model will almost certainly require a high degree of disaggregation. Since analytic solutions can be obtained for only very special types of appreciably simplified economic models, simulation techniques, with their vastly greater capacity for complexity, permit the use of more realistic analogues to describe real economic systems.

The second driving force behind the use of simulation, one that is not unrelated to the first, arises from the need of social scientists in general and of economists in particular to find morally acceptable and scientifically adequate substitutes for the physical scientist's controlled experiments. To the extent that the analogue used in the simulation represents the relevant properties of the economic system under study, experimentation with the ana-



logue can be used to infer the results of analogous experiments with the real economy. The effects in the model of specific changes in the values of particular policy instruments (e.g., taxes, interest rates, price level) can be used to draw at least qualitatively valid inferences concerning the probable effects of analogous changes in the real economic system. Much theoretical analysis in economics is aimed at the study of the probable reactions of an economy to specified exogenous changes, but any economic model that can be studied by analytic techniques must of necessity omit so many obviously relevant considerations that little confidence can be placed in the practical value of the results. Since a simulation study can approximate the economy's behavior and structure considerably more closely, simulation experiments can, at least in principle, lead to conditional predictions of much greater operational significance.

Finally, the mathematical flexibility of simulation permits the use of this tool to gain insights into many phenomena whose intrinsic nature is not at all obvious. It is often possible, for example, to formulate a very detailed quantitative description of a particular process before its essential nature is sufficiently well understood to permit the degree of stylization required for a useful theoretical analysis. Studies of the sensitivity of the results to various changes in assumptions can then be used to disentangle the important from the unimportant features of the problem.

**Past uses of simulation.** The earliest applications of simulation approaches to economics employed physical analogues of a hydraulic or electrical variety. Analogue computers permit the solution of more or less complex linear or nonlinear dynamic systems in which time is treated as a continuous variable. They also enable a visual picture to be gained of adjustment processes. On the other hand, they are much slower than digital computers and can introduce distortions into the results because of physical effects, such as "noise" and friction, which have no conscious economic analogue.

Subsequent economic simulations have tended to rely primarily on digital computers. As the speed and memory capacity of the computers have improved, the economic systems simulated have become increasingly more complex and more elaborate, and the emphasis in simulation has shifted from use as a mathematical tool for solving systems of equations and understanding economic models to use as a device for forecasting and controlling real economies. In addition, these improvements in computer design have permitted the construction of microanalytic models in which ag-

gregate relationships are built up from specifications concerning the behavioral patterns of a large sample of microeconomic units.

As early as 1892, Irving Fisher recommended the use of hydraulic analogies "not only to obtain a clear and analytical picture of the interdependence of the many elements in the causation of prices, but also to employ the mechanism as an instrument of investigation and by it, study some complicated variations which could scarcely be successfully followed without its aid" (1892, p. 44). It was not until 1950, however, that the first hydraulic analogues of economic systems were constructed. Phillips (1950) used machines made of transparent plastic tanks and tubes through which colored water was pumped to depict the Keynesian mechanism of income determination and to indicate how the production, consumption, stocks, and prices of a commodity interact. Electrical analogues were used to study models of inventory determination (Strotz et al. 1957) and to study the business cycle models of Kalecki (Smith & Erdley 1952), Goodwin (Strotz et al. 1953), and others.

The shift to the use of digital computers began in the late 1950s. In a simulation study on an IBM 650, Adelman and Adelman (1959) investigated the dynamic properties of the Klein-Goldberger model of the U.S. economy by extrapolating the exogenous variables and solving the system of 25 nonlinear difference equations for a period of one hundred years. In this process no indications were found of oscillatory behavior in the model. The introduction of random disturbances of a reasonable order of magnitude, however, generated cycles that were remarkably similar to those characterizing the U.S. economy. On the basis of their study, they concluded that (1) the Klein-Goldberger equations may represent good approximations to the behavioral relationships in the real economy; and (2) their results are consistent with the hypothesis that the fluctuations experienced in modern highly developed societies are due to random perturbations.

Duesenberry, Eckstein, and Fromm (1960) constructed a 14-equation quarterly aggregative econometric model of the U.S. economy. Simulation experiments with this model were used to test the vulnerability of the U.S. economy to depressions and to assess the effectiveness of various automatic stabilizers, such as tax declines, increases in transfer payments, and changes in business savings.

A far more detailed and, indeed, the most ambitious macroeconometric simulation effort to date, at least on this side of the iron curtain, is being carried out at the Brookings Institution in Washington, D.C. The simulation is based on a 400-equation

tion quarterly econometric model of the U.S. economy constructed by various experts under the auspices of the Social Science Research Council (Duesenberry et al. 1965). The Brookings model has eight production sectors. For each of the non-government sectors there are equations describing the determinants of fixed investment intentions, fixed investment realizations, new orders, unfilled orders, inventory investment, production worker employment, nonproduction worker employment, production worker average weekly hours, labor compensation rates per hour, real depreciation, capital consumption allowances, indirect business taxes, rental income, interest income, prices, corporate profits, entrepreneurial income, dividends, retained earnings, and inventory valuation adjustment. The remaining expenditure components of the national product are estimated in 5 consumption equations, 11 equations for nonbusiness construction, and several import and export equations. For government, certain nondefense expenditures, especially at the state and local levels, are treated as endogenous variables, while the rest are taken as exogenous. On the income side of the accounts there is a vast array of additional equations for transfer payments, personal income taxes, and other minor items. The model also includes a demographic sector containing labor force, marriage rate, and household formation variables and a financial sector that analyzes the demand for money, the interest rate term structure, and other monetary variables. Finally, aside from a battery of identities, the model also incorporates two matrices: an input-output matrix that translates constant dollar GNP (gross national product) component demands into constant dollar gross product originating by industry, and a matrix that translates industry prices into implicit deflators for GNP components.

The solution to the system is achieved by making use of the model's block recursive structure. First, a block of variables wholly dependent on lagged endogenous variables is solved. Second, an initial simultaneous equation solution is obtained for variables in the "quantity" block (consumption, imports, industry outputs, etc.) using predetermined variables and using prices from the previous period in place of current prices. Third, a simultaneous equation solution is obtained for variables in the "price" block using the predetermined variables and the initial solution from the quantity block. Fourth, the price block solution is used as a new input to the quantity block, and the second and third steps are iterated until each variable changes (from iteration to iteration) by no more

than 0.1 per cent. Fifth, the three blocks are solved recursively.

Simulation experiments show that the system yields predictions both within and beyond the sample period that lie well within the range of accuracy of other models. Simulation studies have also been undertaken to determine the potential impact of personal-income and excise-tax reductions, of government expenditure increases, and of monetary tightness and ease. Stochastic simulations are also being carried out to ascertain the stability properties of the system.

A rather different approach to the estimation of behavioral relationships for the entire socioeconomic system has been developed by Orcutt and his associates (Orcutt et al. 1961). Starting with the observed behavior of microeconomic decision units rather than macroeconomic aggregates, they estimate behavioral relationships for classes of decision units and use these to predict the behavior of the entire system. For this purpose, functions known as operating characteristics are estimated from sample surveys and census data. These functions specify the probabilities of various outcomes, given the values of a set of status variables which indicate the relevant characteristics of the microeconomic units at the start of a period.

Their initial simulation experiments were aimed at forecasting demographic behavior. For example, to estimate the number of deaths occurring during a given month, the following procedure was applied: an operating characteristic,

$$\log P_i(m) = F_1[A_i(m - \frac{1}{2}), R_i, S_i] + (m - m_0)F_2[A_i(m - \frac{1}{2}), R_i, S_i] + F_3(m),$$

was specified. Here,  $P_i(m)$  indicates the probability of the death of individual  $i$  in month  $m$ ,  $A_i(m - \frac{1}{2})$  is the age of individual  $i$  at the start of month  $m$ ,  $R_i$  denotes the race of individual  $i$ , and  $S_i$  denotes the sex of individual  $i$ . The function  $F_1$  is a set of four age-specific mortality tables, one for each race and sex combination, which describe mortality conditions prevailing during the base month  $m_0$ . The function  $F_2$  is used to update each of these mortality tables to month  $m$ . The function  $F_3$  is a cyclic function which accounts for seasonal variations in death rates.

This and other operating characteristics were used in a simulation of the evolution of the U.S. population month by month between 1950 and 1960. A representative sample of the U.S. population consisting of approximately 10,000 individuals was constructed. In each month of the calculation a random drawing determined, in accordance with the probabilities specified by the relevant oper-



ating characteristics, whether each individual in the sample would die, marry, get divorced, or give birth. A regression analysis was then used to compare the actual and predicted demographic changes during the sample period. Close agreement was obtained. This approach is currently being extended to permit analysis of the consumption, saving, and borrowing patterns of U.S. families, the behavior of firms, banks, insurance companies, and labor unions, etc. When all portions of the microanalytic model have been fitted together, simulation runs will be made to explore the consequences of various changes in monetary and fiscal policies.

So far, we have discussed simulation experiments for prediction purposes only. However, the potential of simulation as an aid to policy formulation has not been overlooked by development planners in either free-enterprise economies or socialist countries. A microanalytic simulation model has been formulated at the U.S.S.R. Academy of Sciences in order to guide the management of the production system of the Soviet economy (Cherniak 1963). The model is quite detailed and elaborate. It starts with individual plants at a specific location and ends up with interregional and interindustry tables for the Soviet Union as a whole. In addition, the U.S.S.R. is contemplating the introduction of a 10-step joint man-computer program. The program is to be cyclic, with odd steps being man-operated and even steps being computer-operated. The functions fulfilled by the man-operated steps include the elaboration of the initial basis of the plan, the establishment of the criteria and constraints of the plan, and the evaluation of results. The computer, on the other hand, will perform such tasks as summarizing and balancing the plan, determining optimal solutions to individual models, and simulating the results of planning decisions. This procedure will be tested at a level of aggregation corresponding to an industrial sector of an economic council and will ultimately be extended to the economy as a whole. I have been unable to find any information on the progress of this work more recent than that in Cherniak (1963).

To take an example from the nonsocialist countries, a large-scale interdisciplinary simulation effort was undertaken at Harvard University, at the request of President John F. Kennedy, in order to study the engineering and economic development of a river basin in Pakistan. The major recommendations of the report (White House 1964) are currently being implemented by the Pakistani government, with funds supplied by the U.S. Agency for International Development.

In Venezuela a 250-equation dynamic simulation model has been constructed by Holland, in conjunction with the Venezuelan Planning Agency and the University of Venezuela. This model, which is based on experience gained in an earlier simulation of a stylized underdeveloped economy (Holland & Gillespie 1963), is being used in Venezuela to compare, by means of sensitivity studies, the repercussions of alternative government expenditure programs on government revenues and on induced imports, as well as to check the consistency of the projected rate of growth with the rate of investment and the rate of savings. An interesting macroeconomic simulation of the Ecuadorian economy, in which the society has been disaggregated into four social classes, is currently being carried out by Shubik (1966).

*Monte Carlo studies.* The small sample properties of certain statistical estimators cannot be determined using currently available mathematical analysis alone. In such cases simulation methods are extremely useful. By simulating an economic structure (including stochastic elements) whose parameters are known, one generates samples of "observations" of a given sample size. Each sample is then used to estimate the parameters by several estimation methods. For each method, the distribution of the estimates is compared with the true values of the parameters to determine the properties of the estimator for the given sample size. This approach, known as the Monte Carlo sampling method, has frequently been applied by econometricians in studies of the small sample properties of alternative estimators of simultaneous equation models [see SIMULTANEOUS EQUATION ESTIMATION].

*Limitations and potential.* To a large extent, the very strength of simulation is also its major weakness. As pointed out earlier, any simulation experiment produces no more than a specific numerical case history of the system whose properties are to be investigated. To understand the manner of operation of the system and to disentangle the important effects from the unimportant ones, simulation with different initial conditions and parameters must be undertaken. The results of these sensitivity studies must then be analyzed by the investigator and generalized appropriately. However, if the system is very complex, these tasks may be very difficult indeed. To enable interpretation of results, it is crucial to keep the structure of the simulation model simple and to recognize that, as pointed out in the definition, simulation by no means implies a blind imitation of the actual situation. The simulation model should express

only the logic of the simulated system together with the elements needed for a fruitful synthesis.

Another major problem in the use of simulation is the interaction between theory construction and simulation experiments. In many cases simulation has been used as an alternative to analysis, rather than as a supplementary tool for enriching the realm of what can be investigated by other, more conventional techniques. The inclination to compute rather than think tends to permeate a large number of simulation experiments, in which the investigators tend to be drowned in a mass of output data whose general implications they are unable to analyze. There are, of course, notable exceptions to this phenomenon. In the water-resource project (Dorfman 1964), for example, crude analytic solutions to simplified formulations of a given problem were used to pinpoint the neighborhoods of the solution space in which sensitivity studies to variations in initial conditions and parameters should be undertaken in the more complex over-all system. In another instance, insights gained from a set of simulation experiments were used to formulate theorems which, once the investigator's intuition had been educated by means of the simulation, were proved analytically. Examples of such constructive uses of simulation are unfortunately all too few.

Finally, in many practical applications of simulation to policy and prediction problems, insufficient attention is paid to the correspondence between the system simulated and its analogue. As long as the description incorporated in the simulation model appears to be realistic, the equivalence between the real system and its analogue is often taken for granted, and inferences are drawn from the simulation that supposedly apply to the real economy. Clearly, however, the quality of the input data, the correspondence between the behavior of the outputs of the simulation and the behavior of the analogous variables in the real system, and the sensitivity of results to various features of the stylization should all be investigated before inferences concerning the real world are drawn from simulation experiments.

In summary, simulation techniques have a tremendous potential for both theoretical analysis and policy-oriented investigations. If a model is chosen that constitutes a reasonable representation of the economic interactions of the real world and that is sufficiently simple in its structure to permit intelligent interpretation of the results at the present state of the art, the simulations can be used to acquire a basic understanding of, and a qualitative feeling for, the reactions of a real economy to vari-

ous types of stimuli. The usefulness of the technique will depend crucially, however, upon the validity of the representation of the system to be simulated and upon the quality of the compromise between realism and tractability. Presumably, as the capabilities of both high-speed computers and economists improve, the limitations of simulation will be decreased, and its usefulness for more and more complex problems will be increased.

IRMA ADELMAN

[For additional treatment of the terminology in this article and other relevant material, see *ECONOMETRIC MODELS, AGGREGATE*.]

#### BIBLIOGRAPHY

- General discussions of the simulation of economic processes can be found in Conference on Computer Simulation 1963; Holland & Gillespie 1963; Orcutt et al. 1960; 1961. A comprehensive general bibliography is Shubik 1960. On the application of analogue computers to problems in economic dynamics, see Phillips 1950; 1957; Smith & Erdley 1952; Strotz et al. 1953; 1957. Examples of studies using digital computers are Adelman & Adelman 1959; Cherniak 1963; Dorfman 1964; Duesenberry et al. 1960; 1965; Fromm & Taubman 1966; White House 1964.
- ADELMAN, IRMA; and ADELMAN, FRANK L. (1959) 1965 The Dynamic Properties of the Klein-Goldberger Model. Pages 278-306 in American Economic Association, *Readings in Business Cycles*. Homewood, Ill.: Irwin. → First published in Volume 27 of *Econometrica*.
- CHErNIAK, I. I. 1963 The Electronic Simulation of Information Systems for Central Planning. *Economics of Planning* 3:23-40.
- CONFERENCE ON COMPUTER SIMULATION, UNIVERSITY OF CALIFORNIA, LOS ANGELES, 1961 1963 *Symposium on Simulation Models: Methodology and Applications to the Behavioral Sciences*. Edited by Austin C. Hoggatt and Frederick E. Balderston. Cincinnati, Ohio: South-Western Publishing.
- DORFMAN, ROBERT 1964 Formal Models in the Design of Water Resource Systems. Unpublished manuscript.
- DUSENBERRY, JAMES S.; ECKSTEIN, OTTO; and FROMM, GARY 1960 A Simulation of the United States Economy in Recession. *Econometrica* 28:749-809.
- DUSENBERRY, JAMES S. et al. (editors) 1965 *The Brookings Quarterly Econometric Model of the United States*. Chicago: Rand McNally.
- FISHER, IRVING (1892) 1961 *Mathematical Investigations in the Theory of Value and Prices*. New Haven: Yale Univ. Press.
- FROMM, GARY; and TAUBMAN, PAUL 1966 Policy Simulations With an Econometric Model. Unpublished manuscript.
- HOLLAND, EDWARD P.; and GILLESPIE, ROBERT W. 1963 *Experiments on a Simulated Under-developed Economy: Development Plans and Balance-of-payments Policies*. Cambridge, Mass.: M.I.T. Press.
- ORCUTT, GUY H. et al. 1960 Simulation: A Symposium. *American Economic Review* 50:893-932.
- ORCUTT, GUY H. et al. 1961 *Microanalysis of Socioeconomic Systems: A Simulation Study*. New York: Harper.
- PHILLIPS, A. W. 1950 Mechanical Models in Economic Dynamics. *Economica* New Series 17:283-305.



PHILLIPS, A. W. 1957. Stabilization Policy and the Time-forms of Lagged Responses. *Economic Journal* 67: 265-277.

SHUBIK MARTIN. 1960. Bibliography on Simulation, Gaming, Artificial Intelligence and Allied Topics. *Journal of the American Statistical Association* 55: 736-751.

SHUBIK, MARTIN. 1963. Simulation of Socioeconomic Systems. Part I: An Application to National Income Accounting. In *Public American Economic Council Proceedings for Research in Economics*. Discussion Paper No. 204.

SMITH, O. J. M., and VANDER, H. P. 1959. An Electronic Analogue for an Economic System. *Electrical Engineering* 71: 362-366.

SIDOTS, R. H., CALVERT, J. F., and MOREHOUSE, N. P. 1957. Analogue Computing Techniques Applied to Economics. American Institute of Electrical Engineers, Transactions 66, part 1: 547-563.

SIDOTS, R. H., MANSLEY, J. C., and NARSEN, J. B. JR. 1964. Computers: National House of the Business Council. An International Seminar. *Economist* 91: 411.

WHITE HOUSE. U.S. DEPARTMENT OF INTERIOR. PANEL ON WATER RESOURCES AND VALUES IN WESTERN PARKLANDS. 1964. *Panel on Land and Water Management in the Indian Lands*. Washington: Government Printing Office.

### III POLITICAL PROCESSES

A political game or political simulation is a type of model that represents some aspect of politics. The referent, or "reality," represented by a simulation, may be technique may be some existing, past, or hypothetical system or process. Regardless of the reference system or process depicted in a game or simulation, the model is always a simplified version of the total process. Some political features will be excluded. Those elements of political phenomena represented by the model are reduced in complexity. The model's structure and selected input produce a representation of the process, as well as the nature of possible decisions. These latter elements of models are also applicable to other kinds of models, whether they be natural, physical, or mathematical.

Simulations and games differ from other types of models in that their interested elements are limited to measuring different values as the system or process operates or evolves. In other words, the model does not transform some of the elements of the model. For example, a game model is designed to show the effect of a particular human decision in a game, but there are restrictions to make the game resemble the other rules. Instead of some game, a political game is a model that is designed to show the effect of a particular human decision in a political process. In the case of a political game, the model is designed to show the effect of a particular human decision in a political process.

also occur in the relationships between the players. When the value of one element in the model is changed, related properties can be adjusted accordingly. Because of this ability to handle change and change, Brody has described simulations and games as "operating models" (1963, p. 670).

Games versus simulations. To the users of both techniques, the distinctions between games and simulations are still ambiguous, as some current definitions will illustrate.

Brody (1963) and Guetzkow and his associates (1963) distinguish between (1) machine, or computer, simulations, (2) man, or manual, games, and (3) mixed, or man-computer, simulations. Generally the "machine" referred to is a digital computer, although some simulations involve other types of calculating equipment. The extension of the term "simulation" to cover those models which are a mixture of man and machine does not occur in other definitions. For example, the concept "simulation" is confined by Pool, Abelson, and Fagan (1965) to models that are completely operated by a computer. Rapoport (1964) defines a simulation as a technique in which both the assessment of the situation and the subsequent decisions are made in accordance with explicit and formal rules. When either assessment or decisions are made by human beings without formal rules, the technique is described as a game. When both assessment and decision are made by men, Rapoport stipulates that the device is a scenario. A different interpretation is given by Shubik, who suggests that a game "is invariably concerned with studying human behavior or teaching individuals" (1964, p. 71). On the other hand, a simulation is designated as the representation of a human or human system in which the actual presence of human is not essential because their behavior is part of the game of the simulated environment. The necessary involvement of human in games was apparent in the definitions of Thorrell and Graves (1964) and Dawson (1962). In Dawson's classification all operating models are simulations, but only those which introduce human players are games.

In these definitions of games and simulations, two distinguishing criteria are necessary. The distinctions are differentiated either by the role assigned to human participants or by the role assigned to formal rules of change or transformation. When human participation is the distinguishing feature, a game is described as an operating model in which participants are present and a simulation is a model without players. With rules of transformation as the criterion, the type of model without human







**Domestic and CIAPOM simulations.** The dynamics of them are created by a combination of program activity and quantitative data. In summary, the operations of the international government simulation in January 1966, 1 are computer programmed on a computer. Nine major factors and nine streams enter to contemporary world politics are discussed in the program. To keep it relatively realistic the operator supplies the distribution of the nine major nations along two dimensions: an approximate relative membership (weight factor, trade, and geographical location) from 1966 data; inputs the computer generates the matrix of the international system. In course of each round a matrix is created areas, and other international factors.

The operator then determines as action by one of the major nations against one of the several other nations (as determined by one of the degree of national effort required— from dependent position to self aid). From this information, the computer program determines the proposed cost of the other nations. The effects of the action action and resulting counteractions are then computed for each round (and the international system). With these two parameters the operator may specify another action.

In more more complex computer simulation of international relations, TIMPIL (Technological, Military, and Political) Simulation from 1966 is currently being developed for the Joint War Game Agency of the U.S. Army, Chief of Staff (1966). Based on TIMPIL is the international system political, economic, cultural, and military factors of 45 nations in 20 matrix which are derived from 11 "variables included in the program."

**Comments.** Whereas the Domestic simulation and TIMPIL are computer representations of international systems, *Comments* Paul & Kasser have reported the psychological difficulties in the field which affect their laboratory processing. In the course from the future system representation, it is necessary deal with through policy, and the simulation inputs are messages that present the interaction between international system. The policy factors and the material three factor, however might be adapted for other level of political interaction.

The messages that from the computer are assessed for affect, attitude or feeling of one nation toward another, and national (like responses) directed to the other in response. Each message is given different, or made to depend on the conditions. Domestic system simulation, or programmatic system, that reflect psychological interactions about

the relations and formation of international. For example, some situation is given a message that long war increases pressure international. The timing of international interaction when action are selected by the policy makers for domestic and domestic. The timing program the only interaction the process is in the course of action that are given more about the domestic factor, affect factor. The affect stream represents the future course a perception of how each nation feels toward all the other nations in the world in a way of Comments the action and the war more domestic, but more given about the timing when action is the next phase in the beginning of World War 1. At the end of the simulation work, the differences in the affect measures, and the events by which there were varying national responses differences between the actual historical figures.

**Accelerating election simulation.** The National Election Commission has applied an all-computer simulation in the 1966 and 1968 American presidential elections (1966 to all 1968). The simulation is represented in the computer in the form of units. Every unit was is specified by a number, one of them that reflect geographical region, the size, etc. each geographical cluster political party, and output. The each unit system information is stored in the political characteristics, the example countries is one, part-time, money, and output in various political events. The computer list for the "multi-layers" are based from various system, political, cultural, history, and output (1966). The assumption of various could possible in the course of political activities in the United States, provide a more detailed representation of various that is obtained in any single political factor the simulation can be applied in a particular election, the operators must decide what characteristics and factors are used before. Represented in more of transformation are written to express the impact of the system in different kinds of units.

In the 1966 campaign, a program which contains a political representation procedure in the future election, and given about the government, more something in the course of the election work. The simulation modeling of the system contained, all with the actual results was in 1966, the operators introduced into the computer program, primarily represented the effects of these "transformations" and input, nuclear transformation, and more well face. This time the percentage of the unit in each state directed by history was generated, and the influence with the game system was on.

**Other operating models.** The benefits of the political system simulation national governments and



procedures but do not form an exhaustive list of those techniques currently used in political inquiry. For example, the Simulmatics project is not the only computer simulation of political elections. McPhee (1961), Coleman (1965), and Abelson and Bernstein (1963) have constructed simulations representing not only the election outcome but also the process by which voters form opinions during the campaign. The last model differs from the others by representing a local community referendum.

Other games and simulations range from local politics games (Wood 1964) to games of international balance of power (Kaplan et al. 1960); from simulations of disarmament inspection (Singer & Hinomoto 1965) to those of a developing nation (Scott & Lucas 1966). In addition, other operating models not directly concerned with the study of political phenomena are relevant. Abstract games of logic developed by Layman Allen are used in the study of judicial processes at the Yale Law School. Bargaining games, such as those advanced by Thomas Schelling (1961), and the substantial number of business and management simulations (K. Cohen et al. 1964; American Management Association 1961; Thorelli & Graves 1964) have political applications, as do some of the numerous war games (Giffin 1965).

### Uses of political gaming and simulation

The development and operation of a game or simulation demand time, effort, and finances. Moreover, operating models are beset by such difficulties as the question of their political validity. Despite these problems, games and simulations have been considered to be useful techniques in research, instruction, and policy formation.

**Research.** Many of the research attractions of operating models can be summarized in their ability to permit controlled experimentation in politics. In political reality, determination of the effects of a phenomenon can be severely hindered by confounding influences of potentially related events. Through a carefully designed model, an experimenter can isolate a property and its effects by either deleting the competing mechanisms from the model or holding them constant. Not only can a game or simulation be manipulated, but it can also be subjected to situations without the permanent and possibly harmful consequences that a comparable event might create in the real world.

This ability of the researcher to control the model permits replications and increased access. To establish a generalized pattern, repeated obser-

vations are required. A simulation or game can be assigned the same set of initial conditions and played over and over again, whereas in its reference system the natural occurrence of a situation might appear infrequently and perhaps only once. Moreover, replications of an operating model may reveal a greater range of alternative responses to a situation than could otherwise be identified.

Increased access to the objects of their research also leads students of politics to consider operating models. Players in a political game can be continuously observed; a detailed set of their written and verbal communications can be secured; and they can be asked to respond to interviews, questionnaires, or other test instruments at points selected by the researcher. Computer simulations can be instructed to provide a "print-out" at any specified stage of their operation, thus providing a record for analysis. The diversity of means by which data are readily obtained from such models contrasts with the limited access to the crowded lives of actual policy makers and their often sensitive written materials.

In addition to control, replication, and access, gaming and simulation contribute in at least two ways to theory building. First, the construction of a game or a simulation requires the developer to be explicit about the units and relationships that are to exist in his model and presumably in the political reference system it represents. An essential component of the political process which has been ignored in the design of a game is dramatized for the discerning observer. A computer simulation necessitates even more explication of the relationships among the model's components. An incomplete program may terminate abruptly or result in unintelligible output. Thus, in constructing an operating model a relationship between previously unconnected findings may be discovered. Alternatively, a specific gap in knowledge about a political process may be pinpointed, and hypotheses, required by the model, may be advanced to provide an explanation.

A second value for theory building results from the operating, or dynamic, quality of games and simulations. Static statements of relationships can be transformed into processes which respond to change without the restrictions imposed by such alternatives as linear regression models. According to Abelson and Bernstein (1963, p. 94), "... the static character of statistical models and their reliance on linear assumptions seem to place an upper limit on their potential usefulness. . . . Computer simulation offers a technique for formulating

systemic models, thus promising to meet this need."

It should be recognized that these research assets are not shared equally by all political games and simulations. For example, a computer simulation may not generate the range of potential alternatives to a situation provided by a series of games played with policy experts. On the other hand, control and replication are more easily achieved in a programmed computer simulation than in a political game in which players are given considerable latitude in forming their own patterns of interaction.

**Instruction.** Simulations involving human participants and games are being used in graduate and undergraduate instruction as well as in secondary school teaching. Several evaluations of these techniques for college and university teaching have been reported in political science (e.g., Alger 1963; Cohen 1962) and in other fields (e.g., Thorelli & Graves 1964, especially pp. 25-31; Conference on Business Games 1963). One of these reports (Alger 1963, pp. 185-186) summarizes the frequently cited values of a simulation or game as a teaching aid: "(1) It increases student interest and motivation; (2) It serves as a laboratory for student application and testing of knowledge; (3) It provides insight into the decision-maker's predicament; (4) It offers a miniature but rich model that facilitates comprehension of world realities."

In contrast to the positive evaluations of users like Alger, a critical assessment of the instructional merits of political gaming and simulation has been offered by Cohen (1962). On the basis of a game he conducted, Cohen questioned (1) whether the game increased motivation among students not already challenged by the course; (2) whether the game stimulated interest in more than a narrow segment of the entire subject matter; (3) whether the game misrepresented or neglected critical features of political reality, thereby distorting the image of political phenomena; and (4) whether a comparable investment by both instructor and students in more conventional modes of learning might have offered a larger increment in education.

Systematic educational research has yet to clarify the differing evaluations of games and simulations for instruction. In one educational experiment with a political simulation (Robinson et al. 1966), undergraduate students in each of three courses were divided into two groups controlled for intelligence, grade-point average, and certain personality characteristics. One group augmented its

regular course activities by a weekly discussion of relevant case studies. The other group in each course participated in a continuing simulation for an equal amount of time. Although most measures of student interest did not indicate significant differences between the groups, students evinced more interest in the case studies, while attendance was higher in the simulation group. Similarly, no direct and unmediated difference was found between the two groups on either mastery of facts or principles. To date, the limited educational research on gaming in other fields (e.g., Conference on Business Games 1963) also has produced ambiguous results.

It may be that the more knowledge students have in the subject area of a model or game, the more complex and detailed it must be to provide a satisfactory learning experience. One instructional use of games and simulations that may prove particularly useful with sophisticated students is to involve them in the construction of a model rather than to simply have them assume the role of players.

**Policy formation.** In some respects the application of games and simulations as adjuncts to policy formation is only an extension of their use for research and instruction. Policy makers have employed games or man-machine simulations as training aids for governmental officials in the same manner as management-training programs have incorporated business games. In the Center for International Studies game at the Massachusetts Institute of Technology, "questionnaires returned by participants have revealed that responsible officials . . . place a uniformly high value on the special benefits the games provide, particularly in sharpening their perspectives of alternatives that could arise in crisis situations" (Bloomfield & Whaley 1965, p. 866).

Simulation and gaming research performed under government contract or with more or less direct policy implications has concerned such issues as the systemic consequences of the proliferation of nuclear weapons (Brody 1963), the impact on political decision making of situational characteristics associated with crisis (Hermann 1965), and the potential role of an international police force at various stages of disarmament (Bloomfield & Whaley 1965). An application of simulation research for political campaigns is reflected in the election simulations conducted under contract to the National Committee of the Democratic Party by the Simulmatics Corporation (Pool et al. 1965). One of the major areas of U.S. government-related



activity involving operating models is war gaming and its extension to cover relevant political aspects of national security. Public material on these activities, however, is limited.

### Validity—evaluating correspondence

How are the elements of political reality estimated in a way to make their transmission to an operating model accurate? What are the implications of representing political properties by deterministic or stochastic processes? What are the relative advantages of games compared to simulations? If humans act as participants, how can the motivations of actual political leaders be reproduced? Underlying most of these questions is the problem of validity. Operating models are by definition representations of an existing or potential system or process. They are constructed for the information they can provide about a selected reference system. If in comparison to the performance of the reference system the simulation or game produces spurious results, the model is of little worth. An estimation of the "goodness of fit" or extent of similarity between an operating model and external criteria is, therefore, the central problem of validity.

Different types of criteria or standards can be used for evaluating the correspondence between operating models and their reference system (Hermann 1967). To date, however, the only effort at validity that has been applied to most simulations and games uses no specific criterion of comparison. In this approach, called face validity, the model's realism is based on impressionistic judgments of observers or participants. If participants are used, face validity may concern estimates of their motivation or involvement. Unfortunately, participants can be motivated in a game that incorrectly represents the designated political environment. Furthermore, if the operating model involves the substitution of one property for another, some feature may give the appearance of being quite unreal even though it replicates the performance of the real-world property it replaces.

When specific criteria for correspondence are established, model validity may be determined by comparing events, variables and parameters, or hypotheses. In the first case, the product or outcome of the operating model is compared to the actual events it was intended to replicate (e.g., the correlation between the state-by-state election returns in the 1960 U.S. presidential campaign and those projected by the Simulmatics Corporation's model; Pool et al. 1965). The second approach compares the variables and parameters that consti-

tute a model with the real-world properties they are intended to represent (e.g., a current study uses factor analysis of the core variables of a simulation and the quantifiable real-world indices for which the simulation variables have been substituted; validity judgments are made from comparisons of the factors and of the variables that load on each factor; see Chadwick 1966). The third approach to validity compares the statistical results of a number of hypotheses tested in an operating model with comparable tests of the same hypotheses conducted with data drawn from the reference system (e.g., Zinnes 1966). Confirmation of a variety of hypotheses in both the operating model and in the actual political system increases confidence in the model's validity.

With the exception of face validity, each of the described validity approaches requires that values for specified properties be determined with precision for both the operating model and its reference system. Not only is this procedure often difficult, but in some instances it may not be appropriate. Imagine a political game designed as an aid for policy making whose purpose is to display as many different alternative outcomes to an initial situation as possible. The fact that the subsequent course of actual events leads to only one of those outcomes—or perhaps, none of them—may not reduce the utility of the exercise for increasing the number of alternatives considered by the policy makers. In other instances, an operating model may be consistent with one of the described validity criteria but remain unsatisfactory for some purposes. One illustration is an election simulation in which the winning political party is decided by a stochastic process. With correct probability settings and frequent elections in the model, the simulation's distribution of party victories would closely approximate election outcomes in the reference system. Despite the apparent event validity, the simulation would be undesirable for instruction in election politics. The naive participant might conclude from his experience with the collapsed campaign process that party victory is determined exclusively by chance.

In summary, a variety of validity strategies exist. The appropriate strategy varies from model to model and with the purpose for which the game or simulation was designed. Each operating model must be independently validated. At any given time the confidence in the accuracy with which a game or simulation represents its intended political reference system will be a matter of degree. Evaluation of these operating models for the study of

politics will depend upon the development and application of procedures for measuring the extent to which each model's purposes are fulfilled.

CHARLES F. HERMANN

[See also INTERNATIONAL RELATIONS; STRATEGY.]

#### BIBLIOGRAPHY

- ABELSON, ROBERT P.; and BERNSTEIN, ALEX 1963 A Computer Simulation Model of Community Referendum Controversies. *Public Opinion Quarterly* 27:93-122.
- ALGER, CHADWICK F. 1963 Use of the Inter-Nation Simulation in Undergraduate Teaching. Pages 150-189 in Harold Guetzkow et al., *Simulation in International Relations: Developments for Research and Teaching*. Englewood Cliffs, N.J.: Prentice-Hall.
- AMERICAN MANAGEMENT ASSOCIATION 1961 *Simulation and Gaming: A Symposium*. Report No. 55. New York: The Association.
- BENSON, OLIVER 1961 A Simple Diplomatic Game. Pages 504-511 in James N. Rosenau (editor), *International Politics and Foreign Policy: A Reader in Research and Theory*. New York: Free Press.
- BLOOMFIELD, L. P.; and WHALEY, B. 1965 The Political-Military Exercise: A Progress Report. *Orbis* 8:854-870.
- BRODY, RICHARD A. 1963 Some Systemic Effects of the Spread of Nuclear Weapons Technology: A Study Through Simulation of a Multi-nuclear Future. *Journal of Conflict Resolution* 7:663-753.
- CHADWICK, R. W. 1966 Developments in a Partial Theory of International Behavior: A Test and Extension of Inter-Nation Simulation Theory. Ph.D. dissertation, Northwestern Univ.
- COHEN, BERNARD C. 1962 Political Gaming in the Classroom. *Journal of Politics* 24:367-381.
- COHEN, KALMAN J. et al. 1964 *The Carnegie Tech Management Game: An Experiment in Business Education*. Homewood, Ill.: Irwin.
- COLEMAN, JAMES S. 1964 Collective Decisions. *Sociological Inquiry* 34:166-181.
- COLEMAN, JAMES S. 1965 The Use of Electronic Computers in the Study of Social Organization. *Archives européennes de sociologie* 6:89-107.
- CONFERENCE ON BUSINESS GAMES, TULANE UNIVERSITY, 1961 1963 *Proceedings*. Edited by William R. Dill et al. New Orleans, La.: Tulane Univ., School of Business Administration.
- DAVISON, W. P. 1961 A Public Opinion Game. *Public Opinion Quarterly* 25:210-220.
- DAWSON, RICHARD E. 1962 Simulation in the Social Sciences. Pages 1-15 in Harold Guetzkow (editor), *Simulation in Social Science: Readings*. Englewood Cliffs, N.J.: Prentice-Hall.
- GIFFIN, SIDNEY F. 1965 *The Crisis Game: Simulating International Conflict*. Garden City, N.Y.: Doubleday.
- GOLDHAMER, HERBERT; and SPEIER, HANS (1959) 1961 Some Observations on Political Gaming. Pages 498-503 in James N. Rosenau (editor), *International Politics and Foreign Policy: A Reader in Research and Theory*. New York: Free Press.
- GUETZKOW, HAROLD et al. 1963 *Simulation in International Relations: Developments for Research and Teaching*. Englewood Cliffs, N.J.: Prentice-Hall.
- HERMANN, C. F. 1965 *Crises in Foreign Policy Making: A Simulation of International Politics*. China Lake, Calif.: U.S. Naval Ordnance Test Station, Contract N123 (60S30) 32779A.
- HERMANN, C. F. 1967 Validation Problems in Games and Simulations With Special Reference to Models of International Politics. *Behavioral Science* 12:216-233.
- KAPLAN, MORTON A.; BURNS, ARTHUR L.; and QUANDT, RICHARD E. 1960 Theoretical Analysis of the "Balance of Power." *Behavioral Science* 5:240-252.
- MCPHEE, WILLIAM N. (1961) 1963 Note on a Campaign Simulator. Pages 169-183 in William N. McPhee, *Formal Theories of Mass Behavior*. New York: Free Press.
- POOL, ITHIEL DE SOLA; and KESSLER, A. 1965 The Kaiser, the Tsar, and the Computer: Information Processing in a Crisis. *American Behavioral Scientist* 8, no. 9:31-38.
- POOL, ITHIEL DE SOLA; ABELSON, ROBERT P.; and POPKIN, SAMUEL L. 1965 *Candidates, Issues, and Strategies: A Computer Simulation of the 1960 and 1964 Presidential Elections*. Rev. ed. Cambridge, Mass.: M.I.T. Press.
- RAPOPORT, ANATOL 1964 *Strategy and Conscience*. New York: Harper.
- ROBINSON, JAMES A. et al. 1966 Teaching With Inter-Nation Simulation and Case Studies. *American Political Science Review* 60:53-65.
- SCHELLING, T. C. 1961 Experimental Games and Bargaining Theory. *World Politics* 14:47-68.
- SCOTT, ANDREW M.; and LUCAS, WILLIAM A. 1966 *Simulation and National Development*. New York: Wiley.
- SHUBIK, MARTIN 1964 Game Theory and the Study of Social Behavior: An Introductory Exposition. Pages 3-77 in Martin Shubik (editor), *Game Theory and Related Approaches to Social Behavior: Selections*. New York: Wiley.
- SINGER, J. DAVID; and HINOMOTO, HIROHIDE 1965 Inspecting for Weapons Production: A Modest Computer Simulation. *Journal of Peace Research* 1:18-38.
- THORELLI, HANS B.; and GRAVES, ROBERT L. 1964 *International Operations Simulation: With Comments on Design and Use of Management Games*. New York: Free Press.
- WOOD, R. C. 1964 Smith-Massachusetts Institute of Technology Political Game, Documents 1-5. Unpublished manuscript, Massachusetts Institute of Technology, Department of Political Science.
- U.S. JOINT CHIEFS OF STAFF, JOINT WAR GAMES AGENCY 1965 TEMPER. Volume 1: Orientation Manual. DDC AD 470 375L. Washington: The Agency.
- ZINNES, DINA A. 1966 A Comparison of Hostile Behavior of Decision-makers in Simulate and Historical Data. *World Politics* 18:474-502.

#### SIMULTANEOUS EQUATION ESTIMATION

The distinction between partial and general equilibrium analysis in economic theory is well grounded [see ECONOMIC EQUILIBRIUM]. Early work in econometrics paid inadequate attention to this distinction and overlooked for many years the possibilities of



improving statistical estimates of individual economic relationships by embedding them in models of the economy as a whole [see *ECONOMETRIC MODELS, AGGREGATE*]. The earliest studies in econometrics were concerned with estimating parameters of demand functions, supply functions, production functions, cost functions, and similar tools of economic analysis. The principal statistical procedure used was to estimate the  $\alpha$ 's in the relation

$$y_t = \sum_{i=1}^n \alpha_i x_{it} + u_t, \quad t = 1, 2, \dots, T,$$

using the criterion that  $\sum_{t=1}^T u_t^2$  be minimized. This is the principle of "least squares" applied to a single equation in which  $y_t$  is chosen as the dependent variable and  $x_{1t}, \dots, x_{nt}$  are chosen as the independent variables. The criterion is the minimization of the sum of squared "disturbances" ( $u_t$ ) which are assumed to be unobserved random errors. The estimation of the unknown parameters  $\alpha_i$  is based on the sample of  $T$  observations of  $y_t$  and  $x_{1t}, \dots, x_{nt}$ . This is the usual statistical model and estimation procedure that is used in controlled experimental situations where the set of independent variables consists of selected, fixed variates for the experimental readings on  $y_t$ , the dependent variable. [See *LINEAR HYPOTHESES, article on REGRESSION*.]

However, economics, like most other social sciences, is largely a nonexperimental science, and it is generally not possible to control the values of  $x_{1t}, \dots, x_{nt}$ . The values of the independent variables, like those of the dependent variable, are produced from the general outcome of economic life, and the econometrician is faced with the problem of making statistical inferences from nonexperimental data. This is the basic reason for the use of simultaneous equation methods of estimation in econometrics. In some situations  $x_{1t}, \dots, x_{nt}$  may not be controlled variates, but they may have a one-way causal influence on  $y_t$ . The main point is that least squares yields desirable results only if  $u_t$  is independent of  $x_{1t}, \dots, x_{nt}$ , that is, if  $E(u_t x_{it}) = 0$  for all  $i$  and  $t$ .

**Properties of estimators.** If the  $x_{it}$  are fixed variates, estimators of  $\alpha_i$  obtained by minimizing  $\sum u_t^2$  are *best linear unbiased* estimators. They are linear estimators because, as shown below, they are linear functions of  $y_t$ . An estimator,  $\hat{\alpha}_i$ , of  $\alpha_i$  is called unbiased if

$$E\hat{\alpha}_i = \alpha_i,$$

i.e., if the expected value of the estimator equals the true value. An estimator is best if among all unbiased estimators it has the least variance, i.e., if

$$E(\hat{\alpha}_i - \alpha_i)^2 \leq E(\tilde{\alpha}_i - \alpha_i)^2,$$

where  $\tilde{\alpha}_i$  is any other unbiased estimator. Clearly, the properties of being unbiased and best are desirable ones. These properties are defined without reference to sample size. Two related but weaker properties, which are defined for large samples, are *consistency* and *efficiency*.

An estimator,  $\hat{\alpha}_i$ , is consistent if  $\text{plim } \hat{\alpha}_i = \alpha_i$ , that is, if

$$\lim_{T \rightarrow \infty} P(|\hat{\alpha}_i - \alpha_i| < \epsilon) = 1, \quad \text{for any } \epsilon > 0.$$

This states that the probability that  $\hat{\alpha}_i$  deviates from  $\alpha_i$  by an amount less than any arbitrarily small  $\epsilon$  tends to unity as the sample size  $T$  tends to infinity.

Consider now the class of all consistent estimators that are normally distributed as  $T \rightarrow \infty$ . An *efficient* estimator of  $\alpha_i$  is a consistent estimator whose asymptotic normal distribution has a smaller variance than any other member of this class. [See *ESTIMATION*.]

**Inconsistency of least squares.** The choice of estimators,  $\hat{\alpha}_i$ , such that  $\sum_{t=1}^T (y_t - \sum_{i=1}^n \alpha_i x_{it})^2$  is minimized is formally equivalent to the empirical implementation of the condition that  $E(u_t x_{it}) = 0$ , since the first-order condition for a minimum is

$$\sum_{t=1}^T [(y_t - \sum_{i=1}^n \alpha_i x_{it}) x_{it}] = 0.$$

On the one hand, the  $\alpha_i$  are estimated so as to minimize the residual sum of squares. On the other hand, they are estimated so that the residuals are uncorrelated with  $x_{1t}, \dots, x_{nt}$ . The possible inconsistency of this method is clearly revealed by the latter criterion, for if it is assumed that the  $u_t$  are independent of  $x_{1t}, \dots, x_{nt}$  when they actually are not, the estimators will be inconsistent. This is shown by the formula

$$\text{plim } \hat{\alpha}_i = \alpha_i + \text{plim } \frac{1}{|\mathbf{M}|} \sum_{j=1}^n m_{uj} \mathbf{M}_{ji},$$

where  $\mathbf{M}$  is the moment matrix whose typical element is  $\sum_t x_{it} x_{jt}$ ;  $|\mathbf{M}|$  is the determinant of  $\mathbf{M}$ ;  $m_{uj}$  is  $\sum_t u_t x_{jt}$ ; and  $\mathbf{M}_{ji}$  is the  $j, i$  cofactor of  $\mathbf{M}$ . The inconsistency in the estimator is due to the nonvanishing probability limit of  $m_{uj}$ . In a nonexperimental sample of data, such as that observed as the joint outcome of the uncontrolled simultaneous economic process, we would expect many or all the  $x_{it}$  in a problem to be dependent on  $u_t$ .

**Identifying restrictions.** Since economic models consist of a set of simultaneous equations generating nonexperimental data, the equations of the model must be *identified* prior to statistical estimation. Unless some restrictions are imposed on

specific relationships in a linear system of simultaneous equations, every equation may look alike to the statistician faced with the job of estimating the unknown coefficients. The economist must place a priori restrictions, in advance of statistical estimation, on each of the equations in order to identify them. These restrictions may specify that certain coefficients are known in advance—especially that they are zero, for this is equivalent to excluding the associated variable from an economic relation. Other restrictions may specify linear relationships between the different coefficients.

Consider the generalization of a single equation,

$$y_t = \sum_{i=1}^m \alpha_i z_{it} + u_{it},$$

where  $E(z_{it}u_{it}) = 0$  for all  $i$  and  $t$ , to a whole system,

$$\sum_{j=1}^m \beta_{ij} y_{jt} + \sum_{k=1}^m \gamma_{ik} z_{kt} = u_{it}, \quad i = 1, 2, \dots, n,$$

where  $E(z_{kt}u_{it}) = 0$  for all  $k, i$ , and  $t$ . Every variable enters every equation linearly without restriction, and the statistician has no way of distinguishing one relation from another. Zero restrictions, if imposed, would have the form  $\beta_{rs} = 0$  or  $\gamma_{pq} = 0$ , for some  $r, s, p$ , or  $q$ . In many equations, we may be interested in specifying that sums or differences of variables are economically relevant combinations, i.e., that  $\beta_{rs} = \beta_{ru}$  or that  $\beta_{rs} = -\gamma_{rw}$  or, more generally, that

$$\sum_{s=1}^{m_r} w_{rs} \beta_{rs} + \sum_{s=1}^{m_r} v_{rs} \gamma_{rs} = 0.$$

The last restriction implies that a homogeneous linear combination of parameters in the  $r$ th equation is specified to hold on a priori grounds. The weights  $w_s$  and  $v_s$  are known in advance.

If general linear restrictions are imposed on the equations of a linear system, we may state the following rule: an equation in a linear system is identified if it is not possible to reproduce by linear combination of some or all of the equations in the system an equation having the same statistical form as the equation being estimated.

If the restrictions are of the zero type, a necessary condition for identification of an equation in a linear system of  $n$  equations is that the number of variables excluded from that equation be greater than or equal to  $n - 1$ . A necessary and sufficient condition is that it is possible to form at least one nonvanishing determinant of order  $n - 1$  out of those coefficients, properly arranged, with which the excluded variables appear in the  $n - 1$  other equations (Koopmans et al. 1950).

Criteria for identifiability are stated here for

linear equation systems. A more general treatment in nonlinear systems is given by Fisher (1966). [See STATISTICAL IDENTIFIABILITY.]

### Alternative estimation methods

Assuming that we are dealing with an identified system, let us turn to the problems of estimation. In the system of equations above, the  $y_{jt}$  are *endogenous* or *dependent* variables and are equal in number to the number of equations in the system,  $n$ . The  $z_{kt}$  are *exogenous* variables and are assumed to be *independent* of the disturbances,  $u_{it}$ .

In one of the basic early papers in simultaneous equation estimation (Mann & Wald 1943), it was shown that large-sample theory would, under fairly general conditions, permit lagged values of endogenous variables to be treated like purely exogenous variables as far as consistency in estimation is concerned. Exogenous and lagged endogenous variables are called *predetermined* variables.

Early econometric studies, for example, that of Tinbergen (1939), were concerned with the estimation of a number of individual relationships in which the possible dependence between variables and disturbances was ignored. These studies stimulated Haavelmo (1943) to analyze the consistency problem, for he noted that the Tinbergen model contained many single-equation least squares estimates of equations that were interrelated in the system Tinbergen was constructing, which was intended to be a theoretical framework describing the economy that generated the observations used.

The lack of independence between disturbances and variables can readily be demonstrated. Consider the two-equation system

$$\beta_{11}y_{1t} + \beta_{12}y_{2t} + \sum_{k=1}^m \gamma_{1k}z_{kt} = u_{1t}$$

$$\beta_{21}y_{1t} + \beta_{22}y_{2t} + \sum_{k=1}^m \gamma_{2k}z_{kt} = u_{2t}.$$

The  $z_{kt}$  are by assumption independent of  $u_{1t}$  and  $u_{2t}$ . Some of the  $\gamma$ 's are specified to be zero or are otherwise restricted so that the two equations are identified. Suppose we wish to estimate the first equation. To apply least squares to this equation, we would have to select either  $y_1$  or  $y_2$  as the dependent variable. Suppose we select  $y_1$  and set  $\beta_{11}$  equal to unity. We would then compute the least squares regression of  $y_1$  on  $y_2$  and the  $z_k$  according to the relation

$$y_{1t} = -\beta_{12}y_{2t} - \sum_{k=1}^m \gamma_{1k}z_{kt} + u_{1t},$$

which incorporates all the identifying restrictions on the  $\gamma$ 's.



For this procedure to yield consistent estimators,  $y_{2t}$  must be independent of  $u_{1t}$ . The question is whether the existence of the second equation has any bearing on the independence of  $y_{2t}$  and  $u_{1t}$ . Multiplying the second equation by  $u_{1t}$  and forming expectations, we have

$$E(y_{2t}u_{1t}) = -(\beta_{21}/\beta_{22})E(y_{1t}u_{1t}) + (1/\beta_{22})E(u_{1t}u_{2t}).$$

From the first equation (with  $\beta_{11} = 1$ ), we have

$$E(y_{1t}u_{1t}) = -\beta_{12}E(y_{2t}u_{1t}) + E(u_{1t}^2).$$

Combining these two expressions, we obtain

$$E(y_{2t}u_{1t}) = \frac{E(u_{1t}u_{2t}) - \beta_{12}E(u_{1t}^2)}{\beta_{22} - \beta_{12}\beta_{21}}.$$

In general, this expression does not vanish, and we find that  $y_{2t}$  and  $u_{1t}$  are not independent.

**The maximum likelihood method.** The maximum likelihood method plays a normative role in the estimation of economic relationships, much like that played by perfect competition in economic theory. This method provides consistent and efficient estimators under fairly general conditions. It rests on specific assumptions, and it may be hard to realize all these assumptions in practice or, indeed, to make all the difficult calculations required for solution of the estimation equations.

For the single-equation model, the maximum likelihood method is immediately seen to be equivalent to ordinary least squares estimation for normally distributed disturbances. Let us suppose that  $u_1, \dots, u_T$  are  $T$  independent, normally distributed variables. The  $T$ -element sample has the probability density function

$$p(u_1, \dots, u_T) = \left( \frac{1}{\sqrt{2\pi}\sigma} \right)^T \exp \left( -\frac{1}{2\sigma^2} \sum_{t=1}^T u_t^2 \right).$$

By substitution we can transform this joint density of  $u_1, \dots, u_T$  into a joint density of  $y_1, \dots, y_T$ , given  $x_{11}, \dots, x_{n1}, x_{12}, \dots, x_{n2}, \dots, x_{nT}$ , namely,

$$p(y_1, \dots, y_T | x_{11}, \dots, x_{nT}) = \left( \frac{1}{\sqrt{2\pi}\sigma} \right)^T \exp \left[ -\frac{1}{2\sigma^2} \sum_{t=1}^T \left( y_t - \sum_{i=1}^n \alpha_i x_{it} \right)^2 \right].$$

This function will be denoted as  $L$ , the likelihood function of the sample, and is seen to depend on the unknown parameters  $\alpha_1, \dots, \alpha_n$  and  $\sigma$ . We maximize this function by imposing the following conditions:

$$\frac{\partial \log L}{\partial \alpha_i} = -\frac{1}{\sigma^2} \sum_{t=1}^T \left[ \left( y_t - \sum_{j=1}^n \alpha_j x_{jt} \right) x_{it} \right] = 0, \quad i = 1, \dots, n,$$

$$\frac{\partial \log L}{\partial \sigma} = -\frac{T}{\sigma} + \frac{1}{\sigma^3} \sum_{t=1}^T \left( y_t - \sum_{j=1}^n \alpha_j x_{jt} \right)^2 = 0.$$

These are recognized as the "normal" equations of single-equation least squares theory and the estimation equation for the residual variance—apart from adjustment for degrees of freedom used in estimating  $\sigma^2$ .

In a system of simultaneous equations, we wish to estimate the parameters in

$$\sum_{j=1}^n \beta_{ij} y_{jt} + \sum_{k=1}^m \gamma_{ik} z_{kt} = u_{it}, \quad i = 1, \dots, n.$$

Here we have  $n$  linear simultaneous equations in  $n$  endogenous and  $m$  exogenous variables. The parameters to be estimated are the elements of the  $n \times n$  coefficient matrix

$$\mathbf{B} = (\beta_{ij}),$$

the  $n \times m$  coefficient matrix

$$\mathbf{\Gamma} = (\gamma_{ik}),$$

and the  $n \times n$  variance-covariance matrix

$$\mathbf{\Sigma} = (\sigma_{ij}).$$

The variances and covariances are defined by

$$\sigma_{ij} = E(u_i u_j).$$

A rule of normalization is applied for each equation,

$$\beta_i \Sigma \beta_i' = 1.$$

In practice, one element of  $\beta_i = (\beta_{i1}, \dots, \beta_{in})$  in each equation is singled out and assigned a value of unity.

The likelihood function for the whole system is

$$L = \text{mod } |\mathbf{B}|^T \left( \frac{1}{\sqrt{2\pi}} \right)^{Tn} |\mathbf{\Sigma}|^{-Tn/2} \times \exp \left\{ -\frac{1}{2} \sum_{t=1}^T [(\mathbf{B} \mathbf{y}_t' + \mathbf{\Gamma} \mathbf{z}_t')' \mathbf{\Sigma}^{-1} (\mathbf{B} \mathbf{y}_t' + \mathbf{\Gamma} \mathbf{z}_t')] \right\},$$

where  $\mathbf{y}_t = (y_{1t}, \dots, y_{nt})$ ,  $\mathbf{z}_t = (z_{1t}, \dots, z_{mt})$ ,  $|\mathbf{\Sigma}|$  is the determinant of  $\mathbf{\Sigma}$ , and  $\text{mod } |\mathbf{B}|$  is the absolute value of the determinant of  $\mathbf{B}$  (Koopmans et al. 1950). The matrix  $\mathbf{B}$  enters this expression as the Jacobian of the transformation from the variables  $u_{1t}, \dots, u_{nt}$  to  $y_{1t}, \dots, y_{nt}$ . The problem of maximum likelihood estimation is to maximize  $L$  or  $\log L$  with respect to the elements of  $\mathbf{B}$ ,  $\mathbf{\Gamma}$ , and  $\mathbf{\Sigma}$ . This is especially difficult compared with the similar problem for single equations shown above, because  $\log L$  is highly nonlinear in the unknown parameters, a difficult source of nonlinearity coming from the Jacobian expression  $\text{mod } |\mathbf{B}|$ .

Maximizing  $\log L$  with respect to  $\mathbf{\Sigma}^{-1}$ , we obtain the maximum likelihood estimator of  $\mathbf{\Sigma}$ , which is

$$\hat{\mathbf{\Sigma}} = (\mathbf{B} \mathbf{\Gamma}) \mathbf{M} (\mathbf{B} \mathbf{\Gamma})',$$

where  $\mathbf{M}$  is the moment matrix of the observations, i.e.,

$$M = \frac{1}{T} \begin{bmatrix} y_{11} & \cdots & y_{1T} \\ \vdots & & \vdots \\ y_{n1} & \cdots & y_{nT} \\ \hline z_{11} & \cdots & z_{1T} \\ \vdots & & \vdots \\ z_{m1} & \cdots & z_{mT} \end{bmatrix} \begin{bmatrix} y_{11} & \cdots & y_{n1} & | & z_{11} & \cdots & z_{m1} \\ \vdots & & \vdots & | & \vdots & & \vdots \\ y_{1T} & \cdots & y_{nT} & | & z_{1T} & \cdots & z_{mT} \end{bmatrix}$$

$$= \begin{bmatrix} M_{yy} & M_{yz} \\ M_{zy} & M_{zz} \end{bmatrix}.$$

Substitution of  $\hat{\Sigma}$  into the likelihood function yields the concentrated form of the likelihood function

$$\log L = \text{Const.} + T \log \text{mod } |\mathbf{B}| - (T/2) \log |\hat{\Sigma}|,$$

where Const. is a constant. Hence, we seek estimators of  $\mathbf{B}$  and  $\Gamma$  that maximize

$$\log \text{mod } |\mathbf{B}| - \frac{1}{2} \log |\hat{\Sigma}| \\ = \log \text{mod } |\mathbf{B}| - \frac{1}{2} \log |(\mathbf{B} \Gamma) \mathbf{M} (\mathbf{B} \Gamma)'|.$$

In the single-equation case we minimize the one-element variance expression, written as a function of the  $\alpha_i$ . In the simultaneous equation case, we maximize  $\log \text{mod } |\mathbf{B}| - \frac{1}{2} \log |\hat{\Sigma}|$ , but this can be shown to be equivalent (Chow 1964) to minimization of  $|\hat{\Sigma}|$ , subject to the normalization rule

$$|\mathbf{B} \mathbf{M}_{yy} \mathbf{B}'| = C,$$

where  $C$  is a constant. This normalization is direction normalization, and as long as it is taken into account, scale normalization (such as  $\beta_{ii} = 1$ , cited previously) is arbitrary. Viewed in this way, the method of maximum likelihood applied to a system of equations appears to be a natural generalization of the method of maximum likelihood applied to a single equation, in which case we minimize  $\sigma^2$  subject to a direction-normalization rule.

*Recursive systems.* The concentrated form of the likelihood function shows clearly that a new element is introduced into the estimation process, through the presence of the Jacobian determinant, which makes calculations of the maximizing values of  $\mathbf{B}$  and  $\Gamma$  highly nonlinear. It is therefore worthwhile to search for special situations in which estimation methods simplify at least to the point of being based on linear calculations.

It is evident that the concentrated form of the likelihood function would lend itself to simpler methods of estimating  $\mathbf{B}$  and  $\Gamma$  if  $|\mathbf{B}|$  were a known constant. This would be the case if  $\mathbf{B}$  were triangular, for then, by a scale normalization, we

would have  $\beta_{ii} = 1$  and  $|\mathbf{B}| = 1$ . If  $\mathbf{B}$  is triangular, the system of equations is called a recursive system. We then simply minimize  $|\hat{\Sigma}|$  with respect to the unknown coefficients; this can be looked upon as a generalized variance minimization, an obvious analogue of least squares applied to a single equation.

If, in addition, it can be assumed that  $\Sigma$  is diagonal, maximum likelihood estimators become a series of successive single-equation least squares estimators. Since the matrix  $\mathbf{B}$  is assumed to be triangular, there must be an equation with only one unlagged endogenous variable. This variable (with unit coefficient) is to be regressed on all the predetermined variables in that equation. Next, there will be an equation with one new endogenous variable. This variable is regressed on the preceding endogenous variable and all the predetermined variables in that equation. In the third equation, another new endogenous variable is introduced. It is regressed on the two preceding endogenous variables and all the predetermined variables in that equation, and so on.

If  $\Sigma$  is not diagonal, a statistically consistent procedure would be to use values of the endogenous variables computed from preceding equations in the triangular array instead of using their actual values. Suppose one equation in the system specifies  $y_1$  as a function of certain  $z$ 's. We would regress  $y_1$  on these  $z$ 's and then compute values of  $y_1$  from the relation

$$\hat{y}_{1t} = \sum_{k=1}^m \hat{\gamma}_{1k} z_{kt},$$

where the  $\hat{\gamma}_{1k}$  are the least squares regression estimators of the  $\gamma_{1k}$ . (Some of the  $\gamma_{1k}$  are zero, as a result of the identifying restrictions imposed prior to computing the regression.) Suppose a second equation in the system specifies  $y_2$  as a function of  $y_1$  and certain  $z$ 's. Our next step would be to regress  $y_2$  on  $\hat{y}_1$  and the included  $z$ 's and then compute values of  $y_2$  from the relation

$$\hat{y}_{2t} = \beta_{21} \hat{y}_{1t} + \sum_{k=1}^m \hat{\gamma}_{2k} z_{kt}.$$

The procedure would be continued until all  $n$  equations are estimated.

Methods of dealing with recursive systems have been studied extensively by Wold, and a summary appears in Strotz and Wold (1960). A recursive system without a diagonal  $\Sigma$ -matrix is found in Barger and Klein (1954). One of the most familiar types of recursive systems studied in econometrics is the cobweb model of demand and supply for agricultural products [see BUSINESS CYCLES, article on MATHEMATICAL MODELS].



*Limited-information maximum likelihood.* Another maximum likelihood approach that is widely used is the limited-information maximum likelihood method. It does not hinge on a specific formulation of the model, as do methods for recursive systems; it is a simplified method because it neglects information. As we have seen, identifying restrictions for an equation takes the form of specifying zero values for some parameters or of imposing certain linear relations on some parameters. The term "limited information" refers to the fact that only the restrictions relating to the particular equation (or subset of equations) being estimated are used. Restrictions on other equations in the system are ignored when a particular equation is being estimated.

Let us again consider the linear system

$$B y_i' + \Gamma z_i' = u_i'.$$

These equations make up the *structural form* of the system and are referred to as structural equations. We denote the *reduced form* of this system by

$$y_i' = -B^{-1}\Gamma z_i' + B^{-1}u_i'$$

or

$$y_i' = \Pi z_i' + v_i'.$$

From the reduced form equations select a subset corresponding to the  $n_i$  endogenous variables in a particular structural equation, say equation  $i$ , which is

$$\sum_{j=1}^{n_i} \beta_{ij} y_{jt} + \sum_{k=1}^{m_i} \gamma_{ik} z_{kt} = u_{it}.$$

The summation limit  $m_i$  indicates the number of predetermined variables included in this equation; we have excluded all zero elements in  $\gamma_i$  and indexed the  $z$ 's accordingly. Form the joint distribution of  $v_{1t}, \dots, v_{n_i t}$  over the sample observations and maximize it with respect to the unknown parameters in the  $i$ th structural equation, subject to the restrictions on this equation alone. The restrictions usually take the form

$$0 = \sum_{j=1}^{n_i} \beta_{ij} \pi_{jk}, \quad k = m_i + 1, \dots, m,$$

where there are  $m$  predetermined variables in the whole system; that is, the  $\gamma_{ik}$ ,  $k = m_i + 1, \dots, m$ , are specified to be zero. The estimated coefficients,  $\beta_{ij}$ ,  $\hat{\gamma}_{ik}$ , and  $\hat{\sigma}_i^2$ , obtained from this restricted likelihood maximization are the limited-information estimators. Methods of obtaining these estimators and a study of their properties are given in Anderson and Rubin (1949).

Linear regression calculations are all that are needed in this type of estimation, save for the ex-

traction of a characteristic root of a matrix with dimensionality  $n_i \times n_i$ . A quickly convergent series of iterations involving matrix multiplication leads to the computation of this root and associated vector. The vector obtained, properly normalized by making one coefficient unity, provides estimates of the  $\beta_{ij}$ . The estimates of the  $\gamma_{ik}$  are obtained from

$$\hat{\gamma}_{ik} = - \sum_{j=1}^{n_i} \beta_{ij} \hat{\pi}_{jk}, \quad k = 1, \dots, m_i,$$

where the  $\hat{\pi}_{jk}$  are least squares regression coefficients from the reduced form equations.

It is significant that both full-information and limited-information maximum likelihood estimators are *essentially* unchanged no matter which variable is selected to have a unit coefficient in each equation. That is to say, if we divide through an estimated equation by the coefficient of any endogenous variable, we get a set of coefficients that would have been obtained by applying the estimation methods under the specification that the same variable have the unit coefficient. Full-information and limited-information maximum likelihood estimators are invariant under this type of scale normalization. Other estimators are not.

**Two-stage least squares.** The classical method of least squares multiple regression applied to a single equation that is part of a larger simultaneous system is inconsistent by virtue of the fact that some of the "explanatory" variables in the regression (the variables with unknown coefficients) may not be independent of the error variable. If we can "purify" such variables to make them independent of the error terms, we can apply ordinary least squares methods to the transformed variables. The method of two-stage least squares does this for us.

Let us return to the equation estimated above by limited information. Choose  $y_1$ , say, as the dependent variable, that is, set  $\beta_{11}$  equal to unity. In place of  $y_{2t}, \dots, y_{n_i t}$ , we shall use

$$\hat{y}_{jt} = \sum_{k=1}^{m_i} \hat{\pi}_{jk} z_{kt}, \quad j = 2, \dots, n_i,$$

as explanatory variables. The  $\hat{y}_{jt}$  are *computed* values from the least squares regressions of  $y_j$  on all the  $z_k$  in the system ( $k = 1, \dots, m_i$ ). The coefficients  $\hat{\pi}_{jk}$  are the computed regression coefficients. The regression of  $y_1$  on  $\hat{y}_2, \dots, \hat{y}_{n_i}, z_1, \dots, z_{m_i}$ , provides a two-stage least squares estimator of the single equation. All the equations of a system may be estimated in this way. This can be seen to be a generalization, to systems with nontriangular Jacobians, of the method suggested previously for recursive models in which the variance-covariance matrix of disturbances is not diagonal.

We may write the "normal" equations for these least squares estimators as

$$\begin{bmatrix} \sum_{t=1}^T y_{1t} \hat{y}'_t \\ \sum_{t=1}^T y_{1t} \mathbf{z}'_t \end{bmatrix} = \begin{bmatrix} \mathbf{M}_{\hat{y}\hat{y}} & \mathbf{M}_{\hat{y}\mathbf{z}} \\ \mathbf{M}_{\mathbf{z}\hat{y}} & \mathbf{M}_{\mathbf{z}\mathbf{z}} \end{bmatrix} \begin{bmatrix} -\mathbf{b}' \\ -\mathbf{c}' \end{bmatrix}.$$

In this notation  $\hat{\mathbf{y}}_t$  is the vector of computed values ( $\hat{y}_{2t}, \dots, \hat{y}_{n_t t}$ );  $\mathbf{z}_t$  is the vector ( $z_{1t}, \dots, z_{m_t t}$ );

$$\begin{bmatrix} \mathbf{M}_{\hat{y}\hat{y}} & \mathbf{M}_{\hat{y}\mathbf{z}} \\ \mathbf{M}_{\mathbf{z}\hat{y}} & \mathbf{M}_{\mathbf{z}\mathbf{z}} \end{bmatrix} = \begin{bmatrix} \hat{y}_{21} \cdots \hat{y}_{2T} \\ \vdots \\ \hat{y}_{n_1 1} \cdots \hat{y}_{n_1 T} \\ \hline z_{11} \cdots z_{1T} \\ \vdots \\ z_{m_1 1} \cdots z_{m_1 T} \end{bmatrix} \begin{bmatrix} \hat{y}_{21} \cdots \hat{y}_{n_1 1} & z_{11} \cdots z_{m_1 1} \\ \vdots & \vdots \\ \hat{y}_{2T} \cdots \hat{y}_{n_1 T} & z_{1T} \cdots z_{m_1 T} \end{bmatrix};$$

$\mathbf{b}$  is the estimator of the vector ( $\beta_{12}, \dots, \beta_{1n_1}$ ); and  $\mathbf{c}$  is the estimator of the vector ( $\gamma_{11}, \dots, \gamma_{1m_1}$ ). It should be noted that  $\mathbf{M}_{\hat{y}\mathbf{z}} = \mathbf{M}_{\mathbf{z}\hat{y}} = \mathbf{M}'_{\mathbf{z}\hat{y}}$ . It should be further observed that

$$\mathbf{M}_{\hat{y}\hat{y}} = \mathbf{M}_{\mathbf{y}\mathbf{y}} - \mathbf{M}_{\mathbf{y}\mathbf{z}} \mathbf{M}_{\mathbf{z}\mathbf{z}}^{-1} \mathbf{M}_{\mathbf{z}\mathbf{y}}.$$

In this expression the whole vector  $\mathbf{z}_t = (z_{1t}, \dots, z_{m_t t})$ , which includes all the predetermined variables in the system, is used for the evaluation of the relevant moment matrices.

**k-Class estimators.** Theil (1958) and Basmann (1957), independently, were the first to advocate the method of two-stage least squares. Theil suggested a whole system of estimators, called the *k*-class. He defined these as the solutions to

$$\begin{bmatrix} \sum_{t=1}^T y_{1t} \mathbf{y}'_t - k \sum_{t=1}^T y_{1t} \hat{\mathbf{y}}'_t \\ \sum_{t=1}^T y_{1t} \mathbf{z}'_t \end{bmatrix} = \begin{bmatrix} \mathbf{M}_{\mathbf{y}\mathbf{y}} - k \mathbf{M}_{\mathbf{y}\hat{\mathbf{y}}} \mathbf{M}_{\hat{\mathbf{y}}\hat{\mathbf{y}}}^{-1} \mathbf{M}_{\hat{\mathbf{y}}\mathbf{y}} & \mathbf{M}_{\mathbf{y}\mathbf{z}} \\ \mathbf{M}_{\mathbf{z}\mathbf{y}} & \mathbf{M}_{\mathbf{z}\mathbf{z}} \end{bmatrix} \begin{bmatrix} -\mathbf{b}' \\ -\mathbf{c}' \end{bmatrix}.$$

In this expression  $\hat{\mathbf{y}}_t$  is the vector of residuals computed from the reduced form regressions of  $y_{2t}, \dots, y_{n_t t}$  on all the  $z_{kt}$ . If  $k = 0$ , we have ordinary least squares estimators. If  $k = 1$ , we have two-stage least squares estimators. If  $k = 1 + \lambda$  and  $\lambda$  is the smallest root of the determinantal equation

$$|\mathbf{M}_{\mathbf{y}\mathbf{y}} \mathbf{M}_{\mathbf{z}\mathbf{z}}^{-1} \mathbf{M}_{\mathbf{z}\mathbf{y}} - \mathbf{M}_{\mathbf{y}\mathbf{y}} \mathbf{M}_{\mathbf{y}\hat{\mathbf{y}}}^{-1} \mathbf{M}_{\hat{\mathbf{y}}\mathbf{y}} - \lambda (\mathbf{M}_{\mathbf{y}\mathbf{y}} - \mathbf{M}_{\mathbf{y}\hat{\mathbf{y}}} \mathbf{M}_{\hat{\mathbf{y}}\hat{\mathbf{y}}}^{-1} \mathbf{M}_{\hat{\mathbf{y}}\mathbf{y}})| = 0,$$

we have limited-information maximum likelihood estimators. This is a succinct way of showing the relationships between various single-equation methods of estimation. Of these three members of the *k*-class, ordinary least squares is not consistent; the other two are.

**Three-stage least squares.** Other members of the *k*-class could be cited for special discussion, but a more fruitful generalization of Theil's approach to estimation theory lies in the direction of what he and Zellner (see Zellner & Theil 1962) call three-stage least squares. Chow (1964) has shown that three-stage least squares estimators are better viewed as simultaneous two-stage least squares estimators. Let us denote  $\hat{\mathbf{u}}_t$  as the vector of residuals associated with each equation in a system that has had elements of  $y_{it}$  replaced by  $\hat{y}_{it}$ , that is,  $\hat{\mathbf{u}}_t = (\hat{u}_{1t}, \dots, \hat{u}_{n_t t})$ , where

$$\hat{u}_{it} = \hat{y}_{it} + \sum_{j=1}^n \beta_{ij} \hat{y}_{jt} + \sum_{k=1}^m \gamma_{ik} z_{kt}.$$

In this formulation  $\sum'$  means summation over all values of *j* except *j* = *i*, and it is assumed that the *i*th endogenous variable in the *i*th equation has a unit coefficient. Some elements of  $\beta_{ij}$  and  $\gamma_{ik}$  are zero or otherwise restricted for identification. Define

$$\hat{\Sigma} = \begin{bmatrix} \hat{u}_{11} \cdots \hat{u}_{1T} \\ \vdots \\ \hat{u}_{n_1 1} \cdots \hat{u}_{n_1 T} \end{bmatrix} \begin{bmatrix} \hat{u}_{11} \cdots \hat{u}_{n_1 1} \\ \vdots \\ \hat{u}_{1T} \cdots \hat{u}_{n_1 T} \end{bmatrix}.$$

Minimization of  $|\hat{\Sigma}|$  with respect to  $\beta_{ij}$  and  $\gamma_{ik}$  yields the estimators sought. Thus, we have an extension of the principle of least squares in which the generalized variance is minimized. This is like the principle of maximum likelihood, which also minimizes  $|\Sigma|$ , expressed in terms of the  $\beta_{ij}$  and  $\gamma_{ik}$ , but the direction normalization is different.

Theil and Zellner termed their method three-stage least squares because they first derived two-stage least squares estimators for each single equation in the system. They computed the residual variance for each equation and used these as estimators of the variances of the true (unobserved) random disturbances. They then used Aitken's



generalized method of least squares (1935) to estimate all the equations in the system simultaneously. Aitken's method applies to systems of equations in which the variance-covariance matrix for disturbances is a general known positive definite matrix. Theil and Zellner used the two-stage estimator of this variance-covariance matrix as though it were known. The advantage of this method is that it is of the full-information variety, making use of restrictions on all the equations of the system.

**Other methods.** If the conditions for identification of a single equation are such that there are just enough restrictions to transform linearly and uniquely the reduced form coefficients into the structural coefficients, an indirect least squares method of estimation can be used. Exact identification under zero-type restrictions would enable one to solve

$$0 = \sum_{j=1}^n \beta_{ij} \pi_{jk}, \quad k = m_1 + 1, \dots, m,$$

for a unique set of estimated  $\beta_{ij}$ , apart from scale normalization, given a set of estimated  $\pi_{jk}$ . The latter would be determined from least squares estimators of the reduced forms. Since there are  $n_1 - 1$  of the  $\beta_{ij}$  to be determined, the necessary condition for exact identification here is that  $n_1 - 1 = m - m_1$ .

If there is *underidentification*, i.e., too few a priori restrictions, structural estimation cannot be completed but unrestricted reduced forms can be estimated by the method of least squares. This is the most information that the econometrician can extract when there is lack of identification. Least squares estimators of the reduced form equations are consistent in the underidentified case, but estimates of the structural parameters cannot be made.

**Instrumental variables.** The early discussion of estimation problems in simultaneous equation models contained, on many occasions, applications of a method known as the method of instrumental variables. In estimating the  $i$ th equation of a linear system, i.e.,

$$\sum_{j=1}^{n_1} \beta_{ij} y_{jt} + \sum_{k=1}^{m_1} \gamma_{ik} z_{kt} = u_{it},$$

we may choose  $(n_1 - 1) + m_1$  variables that are independent of  $u_{it}$ . These are known as the instrumental set. Naturally, the exogenous variables in the equation ( $z_{1t}, \dots, z_{m_1 t}$ ) are possible members of this set. In addition, we need  $n_1 - 1$  more instruments from the list of exogenous variables in the system but not in the  $i$ th equation. For this problem let these be denoted as  $x_{2t}, \dots, x_{n_1 t}$ . Since

$E(z_{st} u_{it}) = 0$ ,  $s = 1, \dots, m_1$ , and  $E(x_{rt} u_{it}) = 0$ ,  $r = 2, \dots, n_1$ , we can estimate the unknown parameters from

$$\sum_{j=1}^{n_1} \beta_{ij} \sum_{t=1}^T y_{jt} x_{rt} + \sum_{k=1}^{m_1} \gamma_{ik} \sum_{t=1}^T z_{kt} x_{rt} = 0, \quad r = 2, \dots, n_1,$$

$$\sum_{j=1}^{n_1} \beta_{ij} \sum_{t=1}^T y_{jt} z_{st} + \sum_{k=1}^{m_1} \gamma_{ik} \sum_{t=1}^T z_{kt} z_{st} = 0, \quad s = 1, \dots, m_1.$$

With a scale-normalization rule, such as  $\beta_{i1} = 1$ , we have  $(n_1 - 1) + m_1$  linear equations in the same number of unknown coefficients. In exactly identified models there is no problem in picking the  $x_{rt}$ , for there will always be exactly  $n_1 - 1$   $z$ 's excluded from the  $i$ th equation. The method is then identical with indirect least squares. If  $m - m_1 > n_1 - 1$ , i.e., if there are more exogenous variables outside the  $i$ th equation than there are endogenous variables minus one, we have overidentification, and the number of possible instrumental variables exceeds the minimum needed. In order to avoid the problem of subjective or arbitrary choice among instruments, we turn to the methods of limited information or two-stage least squares. In fact, it is instructive to consider how the method of two-stage least squares resolves this matter. In place of single variables as instruments, it uses linear combinations of them. The computed values

$$\hat{y}_{jt} = \sum_{k=1}^m \hat{\pi}_{jk} z_{kt}, \quad j = 2, \dots, n_1,$$

are the new instruments. We can view the method either as the regression of  $y_i$  on  $\hat{y}_2, \dots, \hat{y}_{n_1}$ ,  $z_1, \dots, z_{m_1}$ , or as instrumental-variable estimators with  $\hat{y}_{2t}, \dots, \hat{y}_{n_1 t}$ ,  $z_{1t}, \dots, z_{m_1 t}$  as the instruments. Both come to the same thing. The method of instrumental variables yields consistent estimators.

**Subgroup averages.** The instrumental-variables method can be applied in different forms. One form was used by Wald (1940) to obtain consistent estimators of a linear relationship between two variables each of which is subject to error. This gives rise to a method that can be used in estimating econometric systems. Wald proposed that the estimator of  $\beta$  in

$$y_t = \alpha + \beta x_t,$$

where  $y_t$  and  $x_t$  are both measured with error, be computed from

$$\hat{\beta} = \frac{\sum_{t=\frac{T}{2}+1}^T y_t - \frac{T}{2} \bar{y}}{\sum_{t=\frac{T}{2}+1}^T x_t - \frac{T}{2} \bar{x}}.$$

He proposed ordering the sample in ascending magnitudes of the variable  $x$ . From two halves of the sample, we determine two sets of mean values of  $y$  and  $x$ . The line joining these means will have a slope given by  $\beta$ . Wald showed the conditions under which these estimates are consistent.

This may be called the method of subgroup averages. It is a very simple method, which may readily be applied to equations with more than two parameters. The sample is split into as many groups as there are unknown parameters to be determined in the equation under consideration. If there are three parameters, for example, the sample may be split into thirds and the parameters estimated from

$$\begin{aligned}\bar{y}_1 &= \alpha + \beta \bar{x}_1 + \gamma \bar{z}_1, \\ \bar{y}_2 &= \alpha + \beta \bar{x}_2 + \gamma \bar{z}_2, \\ \bar{y}_3 &= \alpha + \beta \bar{x}_3 + \gamma \bar{z}_3.\end{aligned}$$

The extension to more parameters is obvious. The method of subgroup averages can be shown to be a form of the instrumental-variables method by an appropriate assignment of values to "dummy" instrumental variables.

Subgroup averages is a very simple method, and it is consistent, but it is not very efficient.

*Simultaneous least squares.* The simultaneous least squares method, suggested by Brown (1960), minimizes the sum of squares of all reduced form disturbances, subject to the parameter restrictions imposed on the system, i.e., it minimizes

$$\sum_{i=1}^T \sum_{t=1}^n v_{it}^2,$$

subject to restrictions. Suppose that the  $v_{it}$  are expressed as functions of the observables and parameters, with all restrictions included; then Brown's method minimizes the sum of the elements on the main diagonal of  $\Sigma_v$ , where  $\Sigma_v$  is the variance-covariance matrix of reduced form disturbances, whereas full-information maximum likelihood minimizes  $|\Sigma_v|$ .

Brown's method has the desirable property of being a full-information method; it is distribution free; it is consistent; but it has the drawback that its results are not invariant under linear transformations of the variables. This drawback can be removed by expressing the reduced form disturbance in standard units

$$v'_{it} = v_{it}/\sigma_{y_i}$$

and minimizing

$$\sum_{i=1}^T \sum_{t=1}^n (v'_{it})^2.$$

## Evaluation of alternative methods

The various approaches to estimation of whole systems of simultaneous equations or individual relationships within such systems are *consistent* except for the single-equation least squares method. If the system is recursive and disturbances are independent between equations, least squares estimators are also consistent. In fact, they are maximum likelihood estimators for normally distributed disturbances. But generally, ordinary least squares estimators are not consistent. They are included in the group of alternatives considered here because they have a time-honored status and because they have minimum variance. In large-sample theory, maximum likelihood estimators of parameters are generally efficient compared with all other estimators. That is why we choose full-information maximum likelihood estimators as norms. They are consistent and efficient. Least squares estimators are minimum-variance estimators if their variances are estimated about estimated (inconsistent) sample means. If their variances are measured about the *true*, or population, values, it is not certain that they are efficient.

Limited-information estimators are less efficient than full-information maximum likelihood estimators. This should be intuitively obvious, since full-information estimators make use of more a priori information; it is proved in Klein (1960). Two-stage least squares estimators have asymptotically the same variance-covariance matrix as limited-information estimators, and three-stage (or simultaneous two-stage) least squares estimators have the same variance-covariance matrix as full-information maximum likelihood estimators. Thus, asymptotically the two kinds of limited-information estimators have the same efficiency, and the two kinds of full-information estimators have the same efficiency. The instrumental-variables or subgroup-averages methods are generally inefficient. Of course, the instrumental-variables method can be pushed to the point where it is the same as two-stage least squares estimation and can thereby gain efficiency.

A desirable aspect of the method of maximum likelihood is that its properties are preserved under a single-valued transformation. Thus, efficient estimators of structural parameters by this method transform into efficient estimators of reduced form parameters. The apparently efficient method of least squares may lose its efficiency under this kind of transformation. In applications of models, we use the reduced form in most cases, not the individual structural equations; therefore the prop-



erties under conditions of transformation from structural to reduced form equations are of extreme importance. Limited-information methods are a form of maximum likelihood methods. Therefore the properties of limited information are preserved under transformation.

To obtain limited-information estimators of the single equation

$$\sum_{j=1}^{n_1} \beta_{ij} y_{jt} + \sum_{k=1}^{m_1} \gamma_{ik} z_{kt} = u_{it},$$

we maximize the joint likelihood of  $v_{1t}, \dots, v_{n_1t}$  in

$$y_{jt} = \sum_{i=1}^m \pi_{ji} z_{it} + v_{jt}, \quad j = 1, \dots, n_1,$$

subject to the restrictions on the  $i$ th equation. In this case only the  $n_1$  reduced forms corresponding to  $y_{1t}, \dots, y_{n_1t}$  are used. It is also possible to simplify calculations, and yet preserve consistency (although at the expense of efficiency), by using fewer than all  $m$  predetermined variables in the reduced forms. In this sense the reduced forms of limited-information estimation are not necessarily unique, and the same endogenous variable appearing in different structural equations of a system may not have the same reduced form expression for each equation estimator. There is yet another sense in which we may derive reduced forms for the method of limited information. After each equation of a complete system has been estimated by the method of limited information, we can derive algebraically a set of reduced forms for the whole system. These would, in fact, be the reduced forms used in forecasting, multiplier analysis, and similar applications of systems. The efficiency property noted above for limited and full information has not been proved for systems of this type of reduced forms, but this has been studied in numerical analysis (see below).

**Ease of computation.** Finally we come to an important practical matter in the comparison of the different methods of estimation—relative ease of computation. Naturally, calculations are simpler and smaller in magnitude for single-equation least squares than for any of the other methods except that of subgroup averages. The method of instrumental variables is of similar computational complexity, but for equations with four or more variables it pays to have the advantage of symmetry in the moment matrices, as is the case with single-equation least squares. This is hardly a consideration with modern electronic computing machines, but it is worth consideration if electric desk machines are being used.

The next-simplest calculations are those for two-

stage least squares. These consist of a repeated application of least squares regression techniques of calculation, but the first regressions computed are of substantial size. There are as many independent variables in the regression as there are predetermined variables in the system, provided there are enough degrees of freedom. Essentially, the method amounts to the calculation of parameters and computed dependent variables in

$$\hat{y}_{jt} = \sum_{k=1}^m \hat{\pi}_{jk} z_{kt}, \quad j = 2, \dots, n_1.$$

Only the "forward" part of this calculation by the standard Gauss–Doolittle method need be made in order to obtain the moment matrix of the  $y_{jt}$ . In the next stage we compute the regression

$$y_{1t} = - \sum_{j=2}^{n_1} b_{1j} \hat{y}_{jt} - \sum_{k=1}^{m_1} c_{1k} z_{kt}.$$

Two important computing problems arise in the first stage. In many systems  $m > T$ ; i.e., there are insufficient degrees of freedom in the sample for evaluation of the reduced forms. We may choose a subset of the  $z_{kt}$ , or we may use principal components of the  $z_{kt}$  (Kloek & Mennes 1960). Systematic and efficient ways of choosing subsets of the  $z_{kt}$  have been developed by taking account of the recursive structure of the model (Fisher 1965). In many economic models  $m$  has been as large as 30 or more, and it is often difficult to make sufficiently accurate evaluation of the reduced form regression equations of this size, given the amount of multicollinearity found in economic data with common trends and cycles. The same procedures used in handling the degrees-of-freedom problem are recommended for getting round the difficulties of multicollinearity. Klein and Nakamura (1962) have shown that multicollinearity problems are less serious in ordinary than in two-stage least squares. They have also shown that these problems increase as we move on to the methods of limited-information and then full-information maximum likelihood.

Limited-information methods require all the computations of two-stage least squares and, in addition, the extraction of a root of an  $n_1 \times n_1$  determinantal equation. The latter calculation can be done in a straightforward manner by iterative matrix multiplication, usually involving fewer than ten iterations.

Both limited information and two-stage least squares are extremely well adapted to modern computers and can be managed without much trouble on electric desk machines.

Three-stage least squares estimators involve the

computation of two-stage estimators for each equation of a system, estimation of a variance-covariance matrix of structural disturbances, and simultaneous solution of a linear equation system of the order of all coefficients in the system. This last step may involve a large number of estimating equations for a model of 30 or more structural equations.

All the previous methods consist of standard linear matrix operations. The extraction of a characteristic root is the only operation that involves nonlinearities, and the desired root can quickly be found by an iterative process of matrix multiplication. Full-information maximum likelihood methods, however, are quite different. The estimation equations are highly nonlinear. For small systems of two, three, or four equations, estimates have been made without much trouble on large computers (Eisenpress 1962) and on desk machines (Chernoff & Divinsky 1953). The problem of finding the maximum of a function as complicated as the joint likelihood function of a system of 15 to 20 or more equations is, however, formidable. Electronic machine programs have been developed for this purpose. The most standardized sets of full-information maximum likelihood calculations are for systems that are fully linear in both parameters and variables. Single-equation methods require linearity only in unknown parameters, and this is a much weaker restriction. Much progress in computation has been made since the first discussion of these econometric methods of estimation, in 1943, but the problem is far from solved, and there is no simple, push-button computation. This is especially true of full-information maximum likelihood.

Efficient programs have recently been developed for calculating full-information maximum likelihood estimates in either linear or nonlinear systems, and these have been applied to models of as many as 15 structural equations, involving more than 60 unknown parameters.

**Generalization of assumptions.** The basis for comparing different estimation methods or for preferring one method over another rests on *asymptotic* theory. The property of consistency is a large-sample property, and the sampling errors used to evaluate efficiency measures are asymptotic formulas. Unfortunately, samples of economic data are frequently not large, especially time series data. The amount of small-sample bias or the small-sample confidence intervals for parameter estimators are not generally known in specific formulas. Constructed numerical experiments, designed according to Monte Carlo methods, have

thrown some light on the small-sample properties. These are reported below.

Another assumption sometimes made for the basic model is that the error terms are mutually independent. We noted above that successive least squares treatment of equations in recursive systems is identical with maximum likelihood estimation when the variance-covariance matrix of structural disturbances is diagonal. This implies mutual independence among contemporaneous disturbances. In a time series model we usually make another assumption, namely, that

$$E(u_{it}, u_{jt'}) = 0, \quad t \neq t', \text{ for all } i, j.$$

The simplest way in which this assumption can be modified is to allow the errors to be related in some linear autoregressive process, such as

$$u_{it} = \sum_{j=1}^n \sum_{k=1}^p \rho_{ijk} u_{jt, t-k} + e_{it}, \quad i = 1, 2, \dots, n,$$

where  $E(e_{it}, e_{jt'}) = 0$  ( $t \neq t'$ , for all  $i, j$ ). In a formal sense joint maximum likelihood estimation of structural parameters and autoregressive coefficients,  $\rho_{ijk}$ , can be laid out in estimation equations, but there are no known instances where these have been solved on a large scale, for the estimation equations are very complicated. For single-equation models or for recursive systems which split into a series of single-equation regressions, the autoregressive parameters of first order have been jointly estimated with structural parameters (Barger & Klein 1954). The principal extensions to larger systems have been in cases where the autoregressive parameters are known *a priori*. Then it is easy to make known autoregressive transformations of the variables and proceed as in the case of independent disturbances. [See TIME SERIES.]

Related to the above two points is the treatment of lagged values of endogenous variables as predetermined variables. The presence of lagged endogenous variables reflects serial correlation among endogenous variables rather than among disturbances. In large samples it can be shown that for purposes of estimation we are justified in treating lagged variables as predetermined, but in small samples we incur bias on this account.

Another assumption regarding the disturbances in simultaneous equation systems is that they are mainly due to neglected or unmeasurable variables that affect or disturb each equation of the model. They are regarded as errors in behavior or technology. From a formal mathematical point of view, they could equally well be regarded as a direct error in observation of the normalized dependent variable in each equation, assuming that the system is



written so that there is a different normalized dependent variable in each equation. There is an implicit assumption that the exogenous variables are measured without error. If we change the model to one in which random errors enter through disturbances to each relation and also through inaccurate observation of each individual variable, we have a more complicated probability scheme, whose estimation properties have not been developed in full generality. This again has been a case for numerical treatment by Monte Carlo methods.

The procedures of estimating simultaneous equation models as though errors are mutually independent when they really are not and as though variables are accurately measured when they really are not are *specification errors*. Other misspecifications of models can occur. For simplicity we assume linearity or, at least, linearity in unknown parameters, but the true model may have a different functional form. Errors may not follow the normal distribution, as we usually assume. [See ERRORS, article on EFFECTS OF ERRORS IN STATISTICAL ASSUMPTIONS.]

Full-information methods are sensitive to specification error because they depend on restrictions imposed throughout an entire system. Single-equation methods depend on a smaller set of restrictions. If an investigator has particular interest in just one equation or in a small sector of the economy, he may incur large specification error by making too superficial a study of the parts of the economy that do not particularly interest him. There is much to be said for using single-equation methods (limited information or two-stage least squares) in situations where one does not have the resources to specify the whole economy adequately.

There are numerous possibilities for specifying models incorrectly. These probably introduce substantial errors in applied work, but they cannot be studied in full generality for there is no particular way of showing all the misspecifications that can occur. We can, however, construct artificial numerical examples of what we believe to be the major specification errors. These are discussed below.

**Sampling experiments.** The effect on estimation methods of using simplified assumptions that are not fully met in real life often cannot be determined by general mathematical analysis. Econometricians have therefore turned to constructing sampling experiments with large-scale computers to test proposed methods of estimation where (1) the sample is small; and (2) there is specification error in the statement of the model, such as

(a) nonzero parameters assumed to be zero, (b) dependent exogenous variables and errors assumed to be independent, (c) imperfectly measured exogenous variables assumed to be perfectly measured, or (d) serially correlated errors assumed to be not serially correlated.

So-called Monte Carlo methods are used to perform the sampling experiments that conceptually underlie sampling error calculations. These sampling experiments are never, in fact, carried out with nonexperimental sources of data, for we cannot relive economic life over and over again; but we can instruct a machine to simulate such an experiment.

Consider a single equation to be estimated by different methods, for example,

$$y_t = \alpha + \beta x_t + u_t, \quad t = 1, 2, \dots, T.$$

Fix  $\alpha$  and  $\beta$  at, say, 3.0 and 0.5, respectively, and set  $T = 30$ . This would correspond to the process

$$y_1 = 3.0 + 0.5x_1 + u_1$$

$$\vdots \quad \vdots \quad \vdots \quad \vdots$$

$$y_{30} = 3.0 + 0.5x_{30} + u_{30}.$$

We also fix the values of the predetermined variables  $x_1, x_2, \dots, x_{30}$  once and for all. We set  $T = 30$  to indicate that we are dealing with a 30-element small sample. A sample of 30 annual observations would be the prototype.

Employing a source of random numbers scaled to have a realistic standard deviation and a zero mean, we draw a set of random numbers  $u_1, \dots, u_{30}$ . We then instruct a machine to use  $u_1, \dots, u_{30}$  and  $x_1, \dots, x_{30}$  to compute  $y_1, \dots, y_{30}$  from the above formulas. From the samples of data,  $y_1, \dots, y_{30}$  and  $x_1, \dots, x_{30}$ , we estimate  $\alpha$  and  $\beta$  by the methods being studied. Let  $\hat{\alpha}$  and  $\hat{\beta}$  be the estimated values. We then draw a new set of random numbers,  $u_1, \dots, u_{30}$ , and repeat the process, using the same values of  $x_1, \dots, x_{30}$ . From many such repetitions, say 100, we have sampling distributions of  $\hat{\alpha}$  and  $\hat{\beta}$ . Means of these distributions, when compared with  $\alpha (= 3.0)$  and  $\beta (= 0.5)$ , indicate bias, if any, and standard deviations or root-mean-square values about 3.0 or 0.5 indicate efficiency. From these sampling distributions we may compare different estimators of  $\alpha$  and  $\beta$ .

What we have said about this simple type of experiment for a single equation can readily be extended to an entire system:

$$By'_t + \Gamma x'_t = \alpha'_t, \quad t = 1, 2, \dots, T.$$

In this case we must start with assumed values of  $B$  and  $\Gamma$ . We choose a  $T$ -element vector of values

for each element of  $\mathbf{z}_t$ , the predetermined variables, and repeated  $T$ -element vectors of values for each element of  $\mathbf{u}_t$ . The random variables are chosen so that their variance-covariance matrix equals some specified set of values. As in the single-equation case,  $T = 30$  or some likely small-sample value. The  $\mathbf{z}_t$  are often chosen in accordance with the values of predetermined variables used in actual models. In practice, Monte Carlo studies of simultaneous equation models have dealt with small systems having only two, three, or four equations.

Two sets of results are of interest from these studies. Estimates of individual elements in  $\mathbf{B}$  and  $\mathbf{\Gamma}$  can be studied and compared for different estimators: estimates of  $\mathbf{B}^{-1}\mathbf{\Gamma}$ , the reduced-form coefficients, can be similarly investigated. In addition, we could form some over-all summary statistic, such as standard error of forecast, for different estimators.

The simplest Monte Carlo experiments have been made to test for small-sample properties alone; they have not introduced measurement errors, serial correlation of disturbances, or other specification errors. Generally speaking, these studies clearly show the bias in single-equation least squares estimates where some of the "independent" variables in the regression calculation are not independent of the random disturbances. Maximum likelihood estimators (full or limited information) show comparatively small bias. The standard deviations of individual parameter estimators are usually smallest for the single-equation least squares method, but this standard deviation is computed about the biased sample mean. If estimated about the true mean, least squares sometimes does not show up well, indicating that bias outweighs efficiency. Full-information maximum likelihood shows up as an efficient method, whether judged in terms of variation about the sample or the true mean. Two-stage least squares estimators appear to have somewhat smaller variance about the true values than do limited-information estimators, and both methods measure up to the efficiency of single-equation least squares methods when variability is measured about the true mean.

Asymptotically, limited-information and two-stage estimators have the same variance-covariance matrices, and they are both inefficient compared with full-information estimators. The Monte Carlo results for small samples are not surprising, although the particular experiments studied give a slight edge to two-stage estimators.

When specification error is introduced, in the form of making an element of  $\mathbf{\Gamma}$  zero in the estimation process when it is actually nonzero in the

population, we find that full-information methods are very sensitive. Both limited-information and two-stage estimators perform better than full-information maximum likelihood. Two-stage estimators are the best among all methods examined in this situation. Limited-information estimators are very sensitive to intercorrelation among predetermined variables.

The principal result for Monte Carlo estimators of reduced form parameters is that transformed single-equation least squares values lose their efficiency properties. Being seriously biased as well, these estimates show a poor over-all rating when used for estimating reduced forms for a system as a whole. Full-information estimators, which are shown in these experiments to be sensitive to specification error, do better in estimating reduced form coefficients than in estimating structural coefficients. Their gain in making use of all the a priori information outweighs the losses due to the misspecification introduced and, in the end, gives them a favorable comparison with ordinary least squares estimators of the reduced form equations that make no use of the a priori information and have no specification error.

If a form of specification error is introduced in a Monte Carlo experiment by having common time trends in elements of  $\mathbf{z}_t$  and  $\mathbf{u}_t$ , so that they are not independent as hypothesized, we find that limited-information estimators are as strongly biased as are ordinary least squares values. If time trend is introduced as an additional variable, however, the limited-information method has small bias.

When observation errors are imposed on the  $\mathbf{z}_t$ , both least squares and limited-information estimators show little change in bias but increases in sampling errors. In this model, it turns out as before that the superior efficiency of least squares estimators of individual structural parameters does not carry over to the estimators of reduced form parameters.

A comprehensive sampling-experiment study of alternative estimators under correctly specified and under misspecified conditions is given in Summers (1965), and Johnston (1963) compares results from several completed Monte Carlo studies. This approach is in its infancy, and further investigations will surely throw new light on the relative merits of different estimation methods.

For some years economists were digesting the modern approach to simultaneous equation estimation introduced by Haavelmo, Mann and Wald, Anderson and Rubin, and Koopmans, Rubin, and Leipnik, and there was a period of little change in



this field. Since the development of the two-stage least squares method by Theil, there have been a number of developments. The methods are undergoing interpretation and revision. New estimators are being suggested, and it is likely that many new results will be forthcoming in the next few decades. Wold (1965) has proposed a method based on iterative least squares that recommends itself by its adaptability to modern computers, its consistency, and its capacity to make use of a priori information on all equations simultaneously and to treat some types of nonlinearity with ease. Also, excellent recent books, by Christ (1966), Goldberger (1964), and Malinvaud (1964), greatly aid instruction in this subject.

LAWRENCE R. KLEIN

[See also LINEAR HYPOTHESES, article on REGRESSION.]

#### BIBLIOGRAPHY

- AITKEN, A. C. 1935 On Least Squares and Linear Combination of Observations. Royal Society of Edinburgh, *Proceedings* 55:42-48.
- ANDERSON, T. W.; and RUBIN, HERMAN 1949 Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations. *Annals of Mathematical Statistics* 20:46-63.
- BARGER, HAROLD; and KLEIN, LAWRENCE R. 1954 A Quarterly Model for the United States Economy. *Journal of the American Statistical Association* 49:413-437.
- BASMAN, R. L. 1957 A Generalized Classical Method of Linear Estimation of Coefficients in a Structural Equation. *Econometrica* 25:77-83.
- BROWN, T. M. 1960 Simultaneous Least Squares: A Distribution Free Method of Equation System Structure Estimation. *International Economic Review* 1: 173-191.
- CHERNOFF, HERMAN; and DIVINSKY, NATHAN 1953 The Computation of Maximum-likelihood Estimates of Linear Structural Equations. Pages 236-269 in Cowles Commission for Research in Economics, *Studies in Econometric Method*. Edited by William C. Hood and Tjalling C. Koopmans. New York: Wiley.
- CHOW, GREGORY C. 1964 A Comparison of Alternative Estimators for Simultaneous Equations. *Econometrica* 32:532-553.
- CHRIST, C. F. 1966 *Econometric Models and Methods*. New York: Wiley.
- EISENPRESS, HARRY 1962 Note on the Computation of Full-information Maximum-likelihood Estimates of Coefficients of a Simultaneous System. *Econometrica* 30:343-348.
- FISHER, FRANKLIN M. 1965 Dynamic Structure and Estimation in Economy-wide Econometric Models. Pages 589-635 in James S. Duesenberry et al., *The Brookings Quarterly Econometric Model of the United States*. Chicago: Rand McNally.
- FISHER, FRANKLIN M. 1966 *The Identification Problem in Econometrics*. New York: McGraw-Hill.
- GOLDBERGER, ARTHUR S. 1964 *Econometric Theory*. New York: Wiley.
- HAAVELMO, TRYGVE 1943 The Statistical Implications of a System of Simultaneous Equations. *Econometrica* 11:1-12.
- JOHNSTON, JOHN 1963 *Econometric Methods*. New York: McGraw-Hill.
- KLEIN, LAWRENCE R. 1960 The Efficiency of Estimation in Econometric Models. Pages 216-232 in Ralph W. Pfouts, *Essays in Economics and Econometrics: A Volume in Honor of Harold Hotelling*. Chapel Hill: Univ. of North Carolina Press.
- KLEIN, LAWRENCE R.; and NAKAMURA, MITSUGU 1962 Singularity in the Equation Systems of Econometrics: Some Aspects of the Problem of Multicollinearity. *International Economic Review* 3:274-299.
- KLOEK, T.; and MENNES, L. B. M. 1960 Simultaneous Equations Estimation Based on Principal Components of Predetermined Variables. *Econometrica* 28:45-61.
- KOOPMANS, TJALLING C.; RUBIN, HERMAN; and LEIPNIK, R. B. (1950) 1958 Measuring the Equation Systems of Dynamic Economics. Pages 53-237 in Tjalling C. Koopmans (editor), *Statistical Inference in Dynamic Economic Models*. Cowles Commission for Research in Economics, Monograph No. 10. New York: Wiley.
- MALINVAUD, EDMOND (1964) 1966 *Statistical Methods of Econometrics*. Chicago: Rand McNally. → First published in French.
- MANN, H. B.; and WALD, ABRAHAM 1943 On the Statistical Treatment of Linear Stochastic Difference Equations. *Econometrica* 11:173-220.
- STOTZ, ROBERT H.; and WOLD, HERMAN 1960 A Triptych on Causal Chain Systems. *Econometrica* 28:417-463.
- SUMMERS, ROBERT 1965 A Capital Intensive Approach to the Small Sample Properties of Various Simultaneous Equation Estimators. *Econometrica* 33:1-41.
- THEIL, HENRI (1959) 1961 *Economic Forecasts and Policy*. 2d ed., rev. Amsterdam: North-Holland Publishing.
- TINBERGEN, JAN 1939 *Statistical Testing of Business-cycle Theories*. Volume 2: Business Cycles in the United States of America: 1919-1932. Geneva: League of Nations, Economic Intelligence Service.
- WALD, ABRAHAM 1940 The Fitting of Straight Lines if Both Variables Are Subject to Error. *Annals of Mathematical Statistics* 11:284-300.
- WOLD, HERMAN 1965 A Fix-point Theorem With Econometric Background. *Arkiv für Matematik* 6:209-240.
- ZELLNER, ARNOLD; and THEIL, HENRI 1962 Three-stage Least Squares: Simultaneous Estimation of Simultaneous Equations. *Econometrica* 30:54-78.

#### SINGLE TAX

See TAXATION, article on PROPERTY TAXES; and the biography of GEORGE.

SISMONDI, J. C. L. SIMONDE DE  
See SIMONDE DE SISMONDI, J. C. L.

#### SITUATIONAL TESTS

See under PERSONALITY MEASUREMENT.

# SIZE DISTRIBUTIONS IN ECONOMICS

The size distributions of certain economic and socioeconomic variables—incomes, wealth, firms, plants, cities, etc.—display remarkably regular patterns. These patterns, or distribution laws, are usually skew, the most important being the Pareto law [see PARETO, *article on CONTRIBUTIONS TO ECONOMICS*] and the log-normal, or Gibrat, law (below). Some disagreement about the patterns actually observed still exists. The empirical distributions often approximate the Gibrat law in the middle ranges of the variables and the Pareto law in the upper ranges. The study of size distributions is concerned with explaining why the observed patterns exist and persist. The answer may be found in the conception of the distribution laws as the steady state equilibria of stochastic processes that describe the underlying economic or demographic forces. A steady state equilibrium is a macroscopic condition that results from the balance of a great number of random microscopic movements proceeding in opposite directions. Thus, in a stationary population a constant age structure is maintained by the annual occurrence of approximately constant numbers of births and deaths—the random events par excellence of human life.

The steady state explanation is evidently inspired by the example of statistical mechanics in which the macroscopic conditions are heat and pressure and the microscopic random movements are performed by the molecules. Characteristically, the steady state is independent of initial conditions, i.e., the initial size distribution. In economic applications this is important because it means that the pattern determined by certain structural constants tends to be re-established after a disturbance is imposed on the process. This will only be the case, however, if the process leading to the steady state is really ergodic, that is, if the influence of initial conditions on the state of the system becomes negligible after a certain time; and it will be relevant in practice only if this time interval is sufficiently short.

The idea that the stable pattern of a distribution might be explained by the interplay of a multitude of small random events was first demonstrated in the case of the normal distribution. The central limit theorem shows that the addition of a great number of small independent random variables yields a variable that is normally distributed, when properly centered and scaled. A stochastic process that leads to a normal distribution is the random walk on a straight line with, for example, a 50 per

cent probability each of a step in one direction and a step in the opposite direction. It is only natural that attempts to explain other distribution patterns should have started from this idea. The first extension was to allow the random walk to proceed on a logarithmic scale. The resulting distribution is log-normal on the natural scale and is known as the log-normal or Gibrat distribution. The basic assumption, in economic terms, is that the chance of a certain proportionate growth or shrinkage is independent of the size already reached—the law of proportionate effect. This law was proposed by J. C. Kapteyn, by Francis Galton, and, later, by Gibrat (1931).

Let size (of towns, firms, incomes) at time  $t$  be denoted by  $Y(t)$ , and let  $\epsilon(t)$  represent a random variable with a certain distribution. We have

$$Y(t) - Y(t-1) = \epsilon(t)Y(t-1),$$

$$Y(t) = (1 + \epsilon(t))Y(t-1)$$

$$= Y(0)(1 + \epsilon(1)) \cdots (1 + \epsilon(t)),$$

where  $Y(0)$  is size at time 0, the initial period. For small time intervals the logarithm of size can be represented as the sum of independent random variables and an initial size which will become negligible as  $t$  grows:

$$\log Y(t) = \log Y(0) + \epsilon(1) + \epsilon(2) \cdots + \epsilon(t).$$

If the random variables  $\epsilon$  are identically distributed with mean  $m$  and variance  $\sigma^2$ , the distribution of  $\log Y(t)$  will be normal with mean  $mt$  and variance  $\sigma^2 t$ .

This random walk corresponds to the process of diffusion in physics which is illustrated by the so-called Brownian movement of particles of dust put into a drop of liquid. Since it implies an ever growing variance, the idea of Gibrat is not itself enough to provide an explanation for a stable distribution. There must be a stabilizing influence to offset the tendency of the variance to increase; indeed, a distinguishing feature of the various theories presently to be reviewed lies in the kind of stabilizer they introduce to offset the diffusion. Two interesting cases may be noted here. One possibility is to modify the law of proportionate effect and assume that the chances of growth decline as size increases. This approach has been taken by Kalecki (1945), who assumes a negative correlation between the size and the jump and obtains a Gibrat law with constant variance. Another possibility is to combine the diffusion process of the random walk with a steady inflow of new, small units (firms, cities, incomes). Some units may continue indefinitely to increase in size, but



their weight will be offset by that of a continuous stream of many new, small entrants, so that both the mean and the variance of the distribution will remain constant. This approach, which leads to the Pareto law, has been taken by Simon (1955).

**Review of various models.** Descriptions of various models will illustrate the methods employed. Models differ with regard to the distribution law explained, the field of application (towns, incomes, etc.), and the type of stochastic process used.

**Champernowne's model.** Champernowne (1953) presents a model that explains the Pareto law for the size distribution of incomes. The stochastic process employed is the so-called Markov chain [see MARKOV CHAINS]. The model is based on a matrix of probabilities of transition from one income class to another in a certain interval of time, say a year. The rows are the income classes of one year, the columns the income classes of the next year. The income classes are chosen in such a way that they are equal on the logarithmic scale (for example, incomes from 1 to 10, from 10 to 100, etc.). The probability of a jump from one income class to the next income class in the course of a year is assumed to be independent of the income from which the jump is made (the law of proportionate effect). The number of income earners is constant.

The number of income earners in income class  $s$  is then determined as follows. The number of incomes in class  $s$  at time  $t + 1$  is

$$f(s, t + 1) = \sum_{u=-n}^1 f(s - u, t)p(u),$$

where  $s$ ,  $u$ , and  $t$  take on integer values,  $p(u)$  is the probability of a jump over  $u$  intervals (i.e., the transition probability), and the size of the jump is constrained to the range  $+1, -n$ . In the steady state equilibrium reached after a sufficiently long time has passed, the action of the transition matrix leaves the distribution unchanged. We then have

$$f(s) = \sum_{u=-n}^1 f(s - u)p(u), \quad s > 0,$$

as  $t \rightarrow \infty$ . This difference equation is solved by putting  $f(s) = z^s$ . The characteristic equation

$$g(z) = \sum_{u=-n}^1 z^{1-u}p(u) - z = 0$$

has two positive real roots, one of which is unity. To assure that the other root will be between 0 and 1, Champernowne introduces the following stability condition:

$$g'(1) = - \sum_{u=-n}^1 up(u) > 0.$$

The relevant solution is  $f(s) = b^s$ ,  $0 < b < 1$ , which gives the number of incomes in income class  $s$ . If the lower bound of this class is the log of the income  $Y_s$ , then the probability of an income exceeding  $Y_s$  is given by

$$\log P(Y_s) = s \log b.$$

Since  $s$  is determined by

$$\log Y_s = sh + \log Y_{\min},$$

where  $h$  is the class interval and  $Y_{\min}$  is the lower boundary of the lowest income class, it follows that

$$\log P(Y_s) = \gamma - \alpha \log Y_s,$$

where the parameters  $\gamma$  and  $\alpha$  are determined by  $b$ ,  $h$ , and  $Y_{\min}$ . This is the Pareto law with Pareto coefficient  $\alpha$ .

Champernowne's stability condition implies that the mathematical expectation of a change in income is negative. This counteracts the diffusion. How can the stability condition be justified on economic grounds? It may be connected with the fact that in this model every income earner who drops out is replaced by a new income earner. Since, in practice, young people have on the average lower and more uniform incomes than old people, the replacement of old income earners by young ones usually means a drop in income. Thus, Champernowne's stability condition, as far as its economic basis is concerned, is very similar to the entry of new, small units that act as a stabilizer in Simon's model.

**Rutherford's model.** Rutherford's model (1955) leads, in his opinion, to the Gibrat law for the size distribution of incomes. Newly entering income earners, assumed to be log-normally distributed at the start, are subject to a random walk and thus to increasing variance during their lifetimes. The process of birth and death of income earners, which is explicitly introduced into the model, acts as the stabilizer.

The distribution of total income is obtained by summing the distributions for all age cohorts that contribute survivors. Rutherford's method is to derive the moments of the distribution by integration over time of the moments for the entrance groups. The distribution is built up "synthetically" from the moments, as it were. In the absence of an analytical solution with a definite distribution law, some disagreement remains about the result.

**Simon's model.** In Simon's model (1955), which leads to what he calls the Yule distribution, the aggregate growth of firms, cities, or incomes is given a priori. The stochastic process apportions this given increment to various units according to

certain rules, which are weakened forms of the law of proportionate effect and rules of new entry. As a consequence of this procedure, there is no possibility of shrinkage of individual units. The given aggregate emphasizes the interdependence of fortunes of different firms (the gain of one is the loss of another)—a point that is neglected in other models, such as that of Steindl (1965). On the other hand, the aggregate is, in reality, not given; it is not independent of the action of the firms, which may increase their total market by advertising, product innovation, and so on.

The process of apportionment may be described as follows. We may conveniently think of populations of cities, so that  $f(n, N)$  is the frequency of cities with  $n$  inhabitants in a total urban population of  $N$ ; to be realistic, we shall assume that a city exceeds a certain minimum number of inhabitants;  $n$  will measure the excess over this minimum, and  $N$  will correspondingly be the sum of these excess populations. An additional urban inhabitant is allocated to a new city with a probability  $\alpha$  and to an existing city, of any size class, with a probability proportionate to the number of (excess) inhabitants in that size class. Then,

$$\begin{aligned} f(n, N+1) - f(n, N) \\ &= \frac{1-\alpha}{N} [(n-1)f(n-1, N) - nf(n, N)], \\ f(1, N+1) - f(1, N) \\ &= \alpha - \frac{1-\alpha}{N} f(1, N). \end{aligned}$$

We assume that there is a steady state solution in which the frequencies of all classes of cities change in the same proportion, that is, in which

$$\frac{f(n, N+1)}{f(n, N)} = \frac{N+1}{N}, \quad \text{for all } n.$$

Using this relation and defining a relative frequency of cities as  $f(n) = f(n, N)/(\alpha N)$ , we obtain from the above equations

$$\begin{aligned} f(n) &= f(n-1) \frac{(1-\alpha)(n-1)}{1 + (1-\alpha)n} \\ f(1) &= \frac{1}{2-\alpha}; \end{aligned}$$

or, setting  $1/(1-\alpha) = \rho$ ,

$$\begin{aligned} f(n) &= \frac{(n-1)(n-2) \cdots 2 \cdot 1}{(n+\rho)(n+\rho-1) \cdots (2+\rho)} f(1) \\ &= \frac{\Gamma(n)\Gamma(\rho+2)}{\Gamma(n+\rho+1)} f(1). \end{aligned}$$

This expression is the Yule distribution. Using a property of the  $\Gamma$ -function, it can be shown that the

Yule distribution asymptotically approaches the Pareto law for large values of  $n$ , that is,  $f(n) \rightarrow n^{-\rho-1}f(1)\Gamma(\rho+2)$  as  $n \rightarrow \infty$ .

This model is applicable to cases in which size is measured by a stock, for example, number of employees of a firm. Simon provides an alternative interpretation of it that applies to flows, such as income and turnover of firms. For example, the total flow of income is given, and each dollar is apportioned to existing and new income earners according to the rules given above.

Using simulation techniques, Ijiri and Simon (1964) show that the pattern of the Yule distribution persists if serial correlation of the growth of individual firms in different periods is assumed. This finding is important because, in reality, growth is often affected by "constitutional" factors, such as financial resources and research done in the past.

*The model of Wold and Whittle.* Wold and Whittle (1957) present a model of the size distribution of wealth in which stability is provided by the turnover of generations, as in Rutherford's model. On the death of a wealth owner, his fortune is divided among his heirs (in equal parts, as a simplification). The diffusion effect is provided by the growth of wealth of living proprietors, which proceeds deterministically at compound interest. The model is shown to lead to a Pareto distribution, the Pareto coefficient depending on the number of heirs to an estate and the ratio of the growth rate of capital to the mortality rate of the wealth owners.

*Steindl's models.* Steindl's models (1965, chapters 2, 3) are designed to explain the size distribution of firms, but they can equally well be applied to the size distribution of cities. The distribution laws obtained are, for large sizes, identical with the Pareto law. Like Rutherford's model, Steindl's models rest on a combination of two stochastic processes. One is a birth-and-death process of the population of cities or firms; the other is a stochastic process of the growth of the city or firm itself.

The way in which the interplay of these two processes brings about the Pareto law can be explained in elementary terms. We start with the size distribution of cities. The number of cities can be explained by a birth process, if we assume that cities do not die. Let us assume that new cities are appearing at a constant rate,  $\epsilon$ , the birth rate of cities. The number of cities increases exponentially, and the age distribution of cities at a given moment of time is

$$(1) \quad R(t) = R(0)e^{-\epsilon t}, \quad \epsilon > 0,$$



where  $t$  is age and  $R(t)$  is the number of cities with age in excess of  $t$ ; in other words,  $R(t)$  is the rank of the town aged  $t + dt$ , and  $R(0)$  is the total number of towns existing at the moment of time considered. The size of the city—its number of inhabitants—increases, on the average, with age. If the rate of births plus immigration,  $\lambda$ , and of deaths plus emigration,  $\mu$ , are constant, we obtain an exponential growth function for the size of the city:

$$(2) \quad n(t) = e^{(\lambda - \mu)t}, \quad \lambda > \mu.$$

Eliminating  $t$  between eqs. (1) and (2), we get

$$(3) \quad \ln R = -\frac{\epsilon}{\lambda - \mu} \ln n + \ln R(0).$$

This is the Pareto law (if  $\epsilon/(\lambda - \mu) > 1$ ), and the Pareto coefficient is seen to be the ratio of the growth rate of the number of cities to the growth rate of a city.

This demonstration, which on the face of it is deterministic in character, can be supplemented by a graphical illustration in which the stochastic features are included. In Figure 1 the distribution of cities according to age is plotted in the vertical ( $\ln R, t$ ) plane. The abscissa shows the age of the city, and the ordinate shows the log of the rank of the city. Each city is thus represented by a dot, and the regression line fitted to these points represents relation (1). In the horizontal ( $t, \ln n$ ) plane, we show the exponential growth of cities with age, as in relation (2). Again each city may be represented by a dot showing age and size. The scatter diagram in the horizontal plane may be regarded as a stochastic transformation of the time variable into the size variable. If the size of each city has been found on the scatter diagram, the cities can be reordered according to size; we then obtain, in the third ( $\ln R, \ln n$ ) plane, the transformed relation (3) between the number of cities (rank) and the size of a city.

If firms are studied, we must take into account the death of firms. We might assume that a firm dies when it ceases to have customers. We can imagine that the age distribution in plane 1 of Figure 1 includes the dead firms; they are automatically eliminated in the transformation to size, being transferred to the size class below one. In the exponential relation (1),  $\epsilon$  must now represent the net rate of growth of the number of firms if the birth of firms is assumed to be a constant ratio of the population.

Figure 1 illustrates how the evolution in time of the number of firms (cities) is mapped onto the cross section of sizes. This process may be compared to sedimentation in geology, where a

historical development is revealed in a cross section of the layers. We can also see how irregularities in the evolution over time will affect the size distribution. If an exceptional spurt of births of new firms occurs at one point of time (after a war, for example), the regression line in plane 1 will be broken and its upper part shifted upward in a parallel fashion. The same thing will happen to the transformed distribution in plane 3.

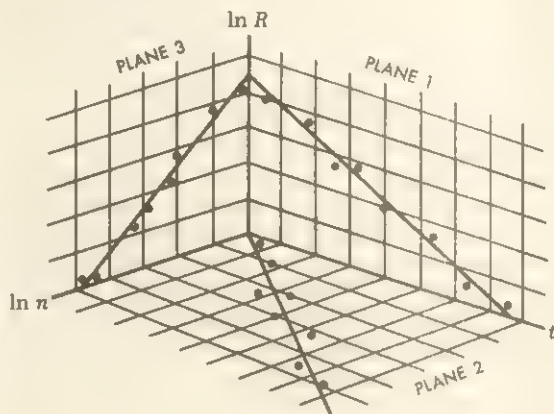


Figure 1

The complete model for firms may be described as follows. The size of a firm is measured by the number of customers attached to it. This is governed by a birth-and-death process. Let us denote by  $o(\Delta t)$  a magnitude that is small in comparison with  $\Delta t$ . There is a chance  $\lambda \Delta t + o(\Delta t)$  of a customer's being acquired and a chance  $\mu \Delta t + o(\Delta t)$  of a customer's being lost in a short period of time,  $\Delta t$ ; multiple births and deaths have a chance of  $o(\Delta t)$ . The probability that a firm has  $n$  or more than  $n$  customers is given by

$$P(n) = \int_0^\infty P(n, t) \tau(t) dt,$$

where  $P(n, t)$  is the probability that a firm of age  $t$  has  $n$  or more than  $n$  customers. The term  $\tau(t)$  is the density of the age distribution of firms, including dead firms; for large  $t$  it is the steady state of a renewal process and is given by  $\tau(t) = ce^{-\epsilon t}$ , where  $\epsilon$  is the net rate of growth of the firm population and  $c$  is a constant. The number of firms with less than one customer,  $P(0, t) - P(1, t)$ , equals the dead firms. The value of  $P(n, t)$  is obtained as the solution of a birth-and-death process for the customers of a firm:

$$P(n, t) = \left(1 - \frac{\mu}{\lambda}\right) \frac{(1 - \alpha')^{n-1}}{(1 - \alpha')^n}, \quad n = 1, 2, \dots,$$

where  $\alpha = e^{-(\lambda - \mu)t}$ ,  $\alpha' = (\mu/\lambda)e^{-(\lambda - \mu)t}$ , and  $\lambda > \mu$ . This expression can be expanded in series and in-

serted in the above integral; this yields, integrating term by term,

$$P(n) = C \sum_{\nu=0}^{\infty} \sum_{k=0}^{n-1} \frac{\left(\frac{-\mu}{\lambda}\right)^{\nu+k} \binom{-n}{\nu} \binom{n-1}{k}}{\nu+k+\omega}$$

$$= C \sum_{\nu=0}^{\infty} \left(-\mu/\lambda\right)^{\nu} \binom{-n}{\nu} B(n, \omega + \nu),$$

where  $\omega = \epsilon/(\lambda - \mu) > 0$  and  $B(n, \omega + \nu)$  is the Beta integral. Hence,

$$P(n) = C \sum_{\nu=0}^{\infty} \left(\frac{\mu}{\lambda}\right)^{\nu} \frac{1}{\nu B(n, \nu)} B(n, \omega + \nu).$$

If  $\mu/\lambda < 1$ , we can neglect the terms with  $\nu$  above a certain value. Thus, if  $n \rightarrow \infty$  and  $\nu$  has a moderate value, we can use the approximations

$$B(n, \omega + \nu) \approx n^{-\omega-\nu} \Gamma(\omega + \nu),$$

$$B(n, \nu) \approx n^{-\nu} \Gamma(\nu);$$

therefore, as  $n \rightarrow \infty$ ,

$$P(n) \rightarrow C'n^{-\omega}.$$

The following features of the solution may be remarked: Since the approximation depends on the value of  $\mu/\lambda$ , which is the mortality of firms of high age, the smaller the mortality of firms, the greater will be the proportion of the distribution that conforms to Pareto's law. The mean of the distribution will be finite if  $\omega > 1$ . This is important in connection with disequilibria, which can arise through changes in  $\lambda$ ,  $\mu$ , and  $\epsilon$ . It can be shown that the Pareto solution applies to the growing firm ( $\lambda > \mu$ , the above case) and, in a modified form, to the shrinking firm ( $\lambda < \mu$ ); but it does not obtain for the stationary firm ( $\lambda = \mu$ ).

The above solution for the distribution according to customers can be shown to be valid also for the distribution according to sales, if firms grow mainly by acquiring more customers and not by getting bigger customers. This is often true in retail trade but not in manufacturing. An alternative model assumes the other extreme—that firms grow only by getting bigger orders. This model is based on the theory of collective risk. The capital of the firm, a continuous variable, is subject to sudden jumps at the instant when orders are executed and to a continuing drain of costs, which is represented deterministically by an exponential decline. The steady state solution obtained from this process is, for large values of capital, identical with the Pareto law; for moderate values, the distribution has a mode and represents, albeit with some complications, a modification of the "first law of Laplace," which was proposed by Fréchet (1939) for income distributions.

**Size as a vector.** It would be natural to measure the size of a firm by a vector, including employment, output, capital, etc., and apply the steady state concept to the joint distribution of several variables. Regression and correlation coefficients obtained in a cross section could then be regarded, like the Pareto coefficient, as characteristics of the steady state. It may be guessed that the growth of the number of firms will have an influence on these parameters as well.

Practically no work has been done in this direction, but it is the only way to clear up the meaning of cross-section data and their relation to time series data and to the theoretical parameters of the underlying stochastic process. The situation in economics is totally unlike that in physics, where the processes are stationary and the ergodic law establishes the identity of time and phase averages. (Only the cosmogony of F. Hoyle, in which the continuous creation of matter offsets the expansion of the universe to establish a steady state of the cosmos, offers a parallel to the growth processes considered above.) The surprise expressed at one time at the difference in estimates of income elasticities from cross-section data and from time series data appears naive in this light because we could only expect them to be equal if the processes generating households, incomes, and consumption were stationary.

But the population of households or the population of firms is not stationary. A cross section of firms shows the growth path of the firm through its different stages of evolution; but the number of firms of a given age depends on the past growth of the total number of firms, and this may influence the regression coefficient. Moreover, the growth path is not unique, because there are several processes superimposed upon one another (growth paths depending on age of firm, age of equipment, age of the management, etc.). For example, the short-run and long-run cost curves are inevitably mixed up in a cross section of firms. [See Cross-section Analysis.]

**How much "stability" and why.** The starting point of the theories here reviewed is the stability of distributions, but stability must not be taken literally. The distributions do change in time, but the change is usually slow. The tail of the distribution of firms or, to a lesser extent, of wealth is composed of very old units, and time must pass before it can be affected by, for example, a change in new entry rates or in growth rates of firms. Thus, the reason for the quasi stability of distributions is that the stock of firms, etc., revolves only slowly. Indirectly this also accounts for the quasi stability of the distribution of incomes, because in-



come is largely determined by wealth or its equivalent in the form of education. An even more enduring influence on the income distribution is the differentiation of skills and professions, which evolves slowly, as a secular process.

The explanations advanced in this article do not exclude the possibility that distribution patterns may change abruptly—for example, as a consequence of taxation, in the case of net incomes; or as a consequence of a big merger movement, in the case of firms.

JOSEF STEINDL

[Directly related are the entries INCOME DISTRIBUTION, article on SIZE; RANK-SIZE RELATIONS.]

#### BIBLIOGRAPHY

- CHAMPERNOWNE, D. G. 1953 A Model of Income Distribution. *Economic Journal* 63:318-351.
- FRÉCHET, MAURICE 1939 Sur les formules de répartition des revenus. International Statistical Institute, *Revue* 7:32-38.
- GIBRAT, ROBERT 1931 *Les inégalités économiques*. Paris: Sirey.
- IJIRI, YUJI; and SIMON, HERBERT A. 1964 Business Firm Growth and Size. *American Economic Review* 54:77-89.
- KALECKI, MICHAEL 1945 On the Gibrat Distribution. *Econometrica* 13:161-170.
- MANSFIELD, EDWIN 1962 Entry, Gibrat's Law, Innovation, and the Growth of Firms. *American Economic Review* 52:1023-1051.
- RUTHERFORD, R. S. G. 1955 Income Distributions: A New Model. *Econometrica* 23:277-294.
- SIMON, HERBERT A. (1955) 1957 On a Class of Skew Distribution Functions. Pages 145-164 in Herbert A. Simon, *Models of Man: Social and Rational*. New York: Wiley. → First published in Volume 42 of *Biometrika*.
- SIMON, HERBERT A.; and BONINI, CHARLES P. 1958 The Size Distribution of Business Firms. *American Economic Review* 48:607-617.
- STEINDL, JOSEF 1965 *Random Processes and the Growth of Firms: A Study of the Pareto Law*. London: Griffin; New York: Hafner.
- WOLD, H. O. A.; and WHITTLE, P. 1957 A Model Exploring the Pareto Distribution of Wealth. *Econometrica* 25:591-595.
- ZIPP, GEORGE K. 1949 *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology*. Reading, Mass.: Addison-Wesley.

## SKILL, ACQUISITION OF

See under LEARNING.

## SKIN SENSES AND KINESTHESIS

The skin senses and kinesthesia (muscle and movement senses) present a difficult problem in sensory physiology—the problem of how information about several energies of the external world are signaled to the brain. Skin receptors are sensitive to thermal and mechanical stimulation, and mechan-

ical stimulation is detected in a variety of ways both by sense organs in cutaneous structures and deep, internal structures such as muscles, tendons, and joints. In varying degrees, deep receptors also detect the state of the tissues or the position of the body and limbs in space.

At one time, it was thought that a single skin sense—so-called common sensibility—detected touch, pressure, warmth, cold, and pain. Beginning in the late nineteenth century, the concept of separate sensory modalities developed, based on the presumed existence of histologically distinct sets of sense organs for each cutaneous sense, of some separation of pathways in the spinal cord, and of specific endings in the cerebral cortex (see Boring 1942). The one concrete experimental datum was a spotlike, or punctiform, distribution of sensitivity, determined by exploring the skin millimeter by millimeter with different energies and finding different psychological responses—touch, warmth, cold—from different spots. This punctiform sensitivity seemed to demand specific anatomical endings. The histological specificity—various encapsulations of the neural terminals themselves—described by anatomists seemed to provide a structural correlate for several psychological and physiological findings. However, the search for a histologically distinctive receptor beneath a sensitive spot proved elusive.

The minimum demanded is that this receptor be a sense organ that responds differentially or differently to various forms of energy. The traditional view is that a sense organ responds differentially to one energy or to one band of an energy spectrum (e.g., warmth and cold or frequencies of light waves) and analyzes an object in contact with the skin in terms of this energy or band of energy. Such a filter function is traditionally expressed by the concept of *adequate stimulus* or, better worded, *differential sensitivity*. Like all filters, the sense organ has a low *threshold* for one form of energy or for one band of frequencies within a given form of energy.

An alternative view holds that sense organs are much alike in their range of sensitivity to a variety of energies but respond differently in some way ("frequency" or "pattern") to each energy, leaving the central nervous system to interpret the different responses and recognize the different energies (Weddell 1961). These points of view are caught in a crosscurrent of contradictory experimental findings.

## Receptors

Since the mid-1950s, the point of view that receptors are alike in their range of sensitivities and

that discrimination of sensations is a joint function of receptor responses and processes in the central nervous system has received considerable support from evidence that there is little morphological differentiation of receptors in hairy skin, although hairy skin indisputably signals different modalities of sensation. While hairless skin (lips, fingertips, etc.) does contain specialized endings, it became gradually accepted, on the basis of Weddell's work (1961), that only two types of endings exist in hairy skin: fibers originating about the roots of hairs and fine, beaded, unmyelinated, or so-called free, nerve endings. Because the same modalities of sensation are experienced from stimulation of hairy and hairless skin, specificity or band-pass characteristics cannot then be attributed to any capsules or discs that are associated with the specialized nerve terminals in hairless skin. Anatomical studies have not disproved the functional specificity of the terminations; whatever specificity exists must, however, be on a subcellular or molecular basis, not visible with the light microscope and conventional stains.

Actually, in slight contradiction of Weddell's broad generalization, there are several types of hair receptors having somewhat different discharge properties, but all are receptors for mechanical stimulation (mechanoreceptors). Also, the conspicuously encapsulated organ—the Pacinian corpuscle—is found in the dermis and subcutaneous tissues. Iggo and Muir (1963) have discovered a morphologically specialized end organ that consists of a thickened epidermis covering several Merkel-type corpuscles, into which penetrate branches of a single thick myelinated axon. This spotlike receptor is also a mechanoreceptor. Deep in the skin and in muscles and tendons, there are encapsulated or otherwise specialized sense organs. Knowledge of sense-organ morphology is in a curious state because certain "classical" formations are now considered artifacts, but at least one *new* type has been discovered.

Another plausible morphologically based generalization, now proved to be invalid, is that all free nerve endings are pain receptors and that their lack of capsule or morphological complexity accounts for the wide variety of energies that can elicit cutaneous pain—namely, mechanical, heat, cold, and chemical stimulation. Passing from the skin to the spinal cord are numerous small axons, measuring about  $1\mu$  in diameter, which conduct nerve impulses at about 1 meter per second and are termed unmyelinated. (Under the light microscopes they appear to lack a surrounding sheath of white fatty material, myelin; electronmicroscopy reveals a thin myelin sheath.) These fibers, termed

C-fibers, conduct impulses which result in painful sensation; so do the smallest fibers with optically visible myelin sheaths. These fibers, called  $\delta$ -fibers, conduct impulses at a considerably faster rate, about 6 meters per second. Because of this double system of fibers, pain can be appreciated, in certain circumstances, as two pulses: "fast" pain and "slow" pain.

It has recently been learned that touch and thermal stimuli excite nerve impulses at intensities of stimulation far below that required to elicit pain. In fact, a large number of the C-fibers can be recruited into responding before the level of stimulation required to elicit pain is reached. Thus, it can no longer be held that the fine unmyelinated axons are concerned exclusively with conducting the impulses that arise within high-threshold end organs and that their central effect results only in the experience of pain. [See PAIN.]

Until recently, knowledge of C-fiber function has been largely inferential, because the small action potentials produced by them have been difficult to record. Douglas, Ritchie, and Straub (1960) caused impulses in the large-diameter, low-threshold axons to collide and, thus, eliminate themselves, leaving only the C-fiber impulses to reach the recording electrode. Activation of C-fibers did not require the strong mechanical or thermal stimulation that elicits pain sensation. In fact, a fall in skin temperature of  $10^\circ\text{C}$ ., less than is required to elicit a painful sensation, activated three-quarters of the C-fibers. Clearly, the C-fibers cannot be concerned solely with pain sensation. This general conclusion is supported by the more direct, classical procedure of "cutting down" a nerve until impulses in only one axon are recorded. This procedure, accomplished by Iggo (1959), showed that C-fibers do respond to more than one form of energy and with weaker stimulation than that required to elicit pain. In fact, they respond to stimuli not much stronger than those required to stimulate the specifically sensitive endings connected with large unmyelinated axons of sensory neurons. C-fiber discharges to a temperature drop as little as  $0.3^\circ\text{C}$ . have been recorded.

Recording from a single axon has been accomplished in man (Hensel & Boman 1960). Receptors connecting with myelinated fibers respond weakly to cold stimulation and strongly to mechanical stimulation. They tend to fire "spontaneously" at natural skin temperature in the absence of known mechanical stimulation.

The one epidermal receptor which is highly structured and is served by a large unmyelinated axon—the touch spot receptor of Iggo and Muir (1963)—discharges in response to cooling. This



receptor is slow in adapting. The muscle spindle, which also adapts slowly, is also sensitive to cooling, but the very rapidly adapting Pacinian corpuscle is not. Iggo (1965) believes that the effect of cooling is a general property of mechanoreceptors which takes time to develop. Postulation of a multispecific ending is not necessary. Whatever the cause, the nervous system has the problem of "decoding" the discharge from such end organs.

In contrast, an ending responding specifically to heat has been discovered. This ending is termed a heat receptor, since it does not respond to strong mechanical and cold stimulation.

**Protopathic and epicritic sensibility.** Since the primitive, fine unmyelinated fibers cannot be associated exclusively with the most primitive and least discriminative sense, the pain modality, and since some myelinated fibers have endings which are thermally, as well as mechanically, sensitive, it is inevitable that thinking should return (Poggio & Mountcastle 1960) to the all but discarded theory of protopathic and epicritic sensibilities, proposed by Henry Head (see Head et al. 1920). The protopathic system was postulated to be sensitive only to extremes of temperature and capable only of gross discriminations, whereas the epicritic was sensitive to light touch and fine gradations of warmth and cold and capable of fine discriminations of pressure and weight. In a sense, a central neural substrate has been furnished by the research just described, assuming always that the afferent neurons described transmit impulses to the brain. Before further consideration of this assumption in the discussion of the spinal pathways of sensation, it should be noted that Head's theory is criticized because quantitative explanations can be given for phenomena which he felt had to be explained in terms of qualitatively different neural mechanisms.

**Deep sensitivity.** Sensory information coming over the deep branches of peripheral nerves is termed deep sensitivity or, specifically, muscle, tendon, and joint sensitivity. Sherrington (1906) introduced the term "proprioception" ("self-sensitivity"), in contrast to "exteroception" ("giving knowledge of the outside world"). "Kinesthesia," or "sense of movement," is favored by some (Rose & Mountcastle 1959), in the belief that the impulses reaching the cerebral cortex come only from joints rather than from muscle and are, therefore, particularly, although not exclusively, excited by movement. The deep receptors lie also in muscle; they record not only active but passive movement and record the posture or position of a limb in space (in neurological jargon, "position sense"). [See SHERRINGTON.]

**Intensity, affect, and place.** So far, this discussion has dealt with the *kind* of stimulus energy that is signaled. The phenomenological counterpart is the *quality* of the sensation, the differences in experience which cause us to give sensations different names such as warmth, cold, touch, or pressure. Sensation has three other aspects which are intriguing neurophysiological puzzles and are of great medical importance: The first of these is *intensity* of stimulation. The second is *affect*, the aspect of a sensory experience which causes us to describe it as pleasant or unpleasant. This phase of sensation blends into the problems of motivation and emotion. The third is *place*, or the spatial-discrimination aspects of cutaneous sensibility.

Intensity is peripherally coded in the frequency of discharge or, perhaps more accurately, in the quantity of impulses—that is, frequency times number of responding fibers. It is no longer likely that affect can be fiber-coded in the periphery—for example, that pain can be identified with C-fibers as a group. The problem becomes one of the destination of the impulse within the brain.

Spatial discrimination involves the ability to locate on the skin surface the point of stimulation (topognosis) and an allied function, so-called two-point discrimination, measured as the smallest distance at which two points applied to the skin are sensed as two rather than one—a function comparable to visual acuity. Errors of localization and fineness of point discrimination are small on the skin surfaces used to explore objects—fingertips, lips, and tongue—and poorly developed on the back and proximal segments of the limbs. Acuity is related to the density of the receptors and to the number of sensory axons supplying the region.

In spatial discriminations, cutaneous and deep receptors combine to provide information about the form of an object palpated and, with the sensations of temperature, weight, etc., enable us to recognize common objects by palpation (stereognosis). Although it is customary to speak loosely of these as sensory processes, technically they are perceptions and judgments. A considerable evolution of the brain has been required to add to the mere recognition of touch, cold, warmth, or pain the abilities to discriminate their fine degrees of intensity, to locate the point on the skin stimulated, and to appreciate the spatial aspects of objects. [For a detailed discussion of intensity and the spatial and affective aspects of sensation, see Ruch & Patton 1965; see also PERCEPTION, article on DEPTH PERCEPTION.]

**Convergence of axons.** Another threat to the concept of specific receptors connected by separate

pathways to the cerebral cortex is the discovery that afferent axons that are apparently specific in the periphery for different kinds of energy converge on the same second-order neuron on the spinal cord. This occurs with axons that terminate on neurons in the posterior horn but not for all second-order neurons on which the posterior-column fibers terminate. Specificity (filtering) and the resulting information would be lost if the transmitting channel had a wide-band pass, unless the streams of impulses from the two narrow-band-pass sense organs were coded differently in the second and subsequent neurons. It has not yet been demonstrated that the single secondary neuron in which the nerve impulses are detected gives rise to ascending sensory fibers. However, similar nonspecific behavior is found in some third-order neurons and some cortical neurons, as will be discussed below.

The simple alternative to specific sensory end organs is frequency coding. A sense organ may fire at one rate when stimulated by touch and at another rate when stimulated by pressure. Alternatively, two specific end organs may, through their axons, fire a second-order neuron at two different rates. The difficulty with either concept is that the frequency code is, so to speak, needed to signal the intensity of a stimulus, as can be seen when the intimate nature of sense-organ excitation is examined.

**Receptor potentials and nerve impulses.** The mechanoreceptor is a transducer that converts non-electrical energy into an electrochemical phenomenon, the generator, or receptor, potential. This potential, in turn, sets up action potentials in the parent axon which are brief, self-propagating, and independent, in size and velocity, of the strength of the stimulus or of the generator potential. By contrast, the magnitude of the generator potential is related in logarithmic proportion to the magnitude of the physical stimulus. Further, the generator potential is stationary, decreasing rapidly with distance along the axon; in response to a sustained stimulus, it tends to be prolonged, although not usually sustained at its initial level. The propagated all-or-none action potentials (nerve impulses) are always intermittent. When the generator potential reaches a certain value (a threshold), the axon is discharged and cannot again be fired until certain changes underlying self-propagation are reversed. [See NERVOUS SYSTEM, especially the articles on STRUCTURE AND FUNCTION OF THE BRAIN and ELECTROENCEPHALOGRAPHY.]

In some receptors, the generator potential dies out so rapidly that only one or two impulses are

set up. Such phasic receptors can signal intensity by their proximity to the stimulus or by differing axon thresholds, which permit stronger and stronger stimuli to involve more and more axons. Such a recruitment will, through convergence at the second-order neuron, mean that more impulses per second reach the brain.

In the tonic type of sensory neuron, the generator potential persists as long as there is a stimulus, but as mentioned, the potential tends to diminish from its initial level to an approximate plateau. So long as this potential exceeds axon threshold, impulses will continue to be generated. However, as the generator potential diminishes, so does the rate of discharge. Since continuance of a stimulus is accompanied by a psychological decrease in intensity (adaptation), the declining frequency of sensory-axon discharge is consistent with the supposition that frequency is the way in which stimulus intensity is coded. More directly, if different plateau levels of the generator potential are induced by different magnitudes of stimulation, the frequency of discharge is *directly and linearly* proportional to the generator potential. The over-all logarithmic relationship between intensity of stimulus and frequency of nerve impulses, characteristic of many receptors, lies in the logarithmic relationship between the magnitude of the physical stimulus and the generator potential. From an engineering point of view, sense organs are poor transducers in respect to linearity, although good in respect to reducing a wide range of intensities to the somewhat limited frequency capacity of nerve axons (600 pulses per second being the upper limit).

### Central sensory processes

The classical somatosensory (skin senses and kinesthesia) pathway in man, monkeys, and apes is made up of systems originating in both the posterior and anterolateral columns of the spinal cord. All sensory pathways are synaptically interrupted. Anterolateral-column pathways first synapse near the point of entry into the spinal cord. Posterior-column pathways synapse for the first time in the lowest region of the brain stem, the medulla.

In both pathways, crossing occurs after the first synapse, so that the left side of the brain is concerned with the right side of the body, as is true of vision and motor function. In man and monkey, the second-order neurons of both sensory pathways synapse in the ventroposterior portion of the thalamus with third-order neurons which project to the cortical areas. The classical somatosensory pathway is conducted by the medial and spinothalamic lemnisci to the posterolateral point of the ventral



nucleus, where it ends in a topographically organized fashion, so that the body surface is projected upon this nucleus (for the detail of this pathway, see Ruch & Patton 1965).

Some pathways, especially those of the anterolateral column, have more than three neurons. In the cat, for example, the *spinothalamic tract*, in the literal sense of the term ("one neuron from cord to thalamus"), does not exist; all of the second-order myelinated fibers synapse before they reach the thalamus. This distinction can be conveyed by speaking of spinothalamic tracts and spinothalamic chains or systems. With the present techniques for studying experimental degeneration in myelinated fibers by staining the axon rather than the disintegrating myelin sheath (and, of course, unmyelinated fibers), the extensiveness of the multisynaptic pathways is just becoming appreciated. In brief, at various levels of the brain stem, many ascending axons give off branches to or terminate on neurons in the central gray matter and reticular substance (a deep-lying meshwork of fibers containing clusters of cell bodies). From these levels, neuron chains proceed to certain thalamic areas which in turn connect with the cerebral cortex. Since the C-fiber system is stimulated peripherally by other than noxious stimuli, the multineuron, fine-fiber ascending system is potentially concerned with all forms of cutaneous sensation.

Two questions can be asked concerning the transmission of sensory information in the spinal cord through the thalamus to the cerebral cortex: First, to what degree is the energy specificity of the sense organ preserved at the first synaptic junction? Second, to what degree do the second-order and third-order neurons receive impulses from a restricted area of the skin? The answer to the latter is indicative of, but not a measure of, the preservation of locus of stimulus in ascending pathways.

Wall (1961), in recording from single second-order neurons of the posterior horn (not to be confused with the posterior column) of the spinal cord, has discovered that they tend to be fired by a variety of stimuli. The axons of these neurons ascend in the anterolateral columns, and the impression is gained that in some anterolateral systems the degree of fiber coding attained in the periphery is largely sacrificed at the first synapse. The kind of stimulation must, therefore, be coded in some other dimension of the stream of afferent impulses. It should be mentioned that reflex connections, as well as ascending axons, originate in the posterior horn and that many anterolateral and posterolateral ascending pathways beside the spinothalamic tract have cells of origin in this region.

It seems, then, that in some ascending pathways, specificity in relation to stimulus energy has been lost in the transfer to secondary axons; either the information itself is lost or else it is coded in some other fashion. Temporal patterning is suggested as the means of coding (Wall 1961).

**Posterior-column pathways.** The behavior of the sensory pathways which ascend the posterior columns of the spinal cord without synapsing and without crossing until they reach the medulla is quite different. In the medulla, the second-order neurons, after crossing, give rise to the medial lemniscus, which ascends the brain stem to the thalamus. The transmission of information which conveys the site of stimulation on the skin is considered by most neurophysiologists to be a product of a topographical organization of axons within ascending pathways that is preserved throughout the thalamic relay nuclei and the thalamocortical projections. The result is a kind of "isomorphism" between the peripheral array of skin receptors and the array of neurons making up the cerebral cortex. This organization must not be thought of in terms so crude as a single receptor being connected with a single cortical cell, an adjacent receptor with an adjacent cortical cell, etc. Such "structural canalization" of pathways seems to be almost deliberately avoided, even in systems—such as the visual—which have a high degree of acuity and spatial discrimination.

**Receptive fields.** There is much interdigitation of free nerve endings in the skin which would seem unnecessary unless some biological value is attained. On the other hand, the number of channels and the degree of their "insulation" from each other (i.e., lack of convergence) at synaptic levels are highest for the points of greatest acuity and spatial discrimination (e.g., the fovea of the retina and the tips of the fingers). One index of topographical organization and potential acuity is the "receptive field." The receptive field of a single axon or cell body is that area of skin from which the unit can be discharged (or, in the case of the eye, the area of the visual field in which the unit can be discharged or inhibited). Thus, each peripheral axon, secondary neuron, and cortical cell has a receptive field.

The receptive fields of the second-order neurons of the posterior-column and the anterolateral-column pathways contrast sharply; the former are measured in millimeters, and the latter, in centimeters. Although large, the receptive fields of second-order neurons on the anterolateral pathways are not splotchy but homogeneous and are often round or elliptical, indicating that many peripheral

axons with adjacent or overlapping fields on the skin have converged in an orderly fashion.

Another finding, which seems to pertain to even the posterior column system, is that receptive fields are larger at the cortical level than at the level of the second-order neuron. This rule also holds in the visual system, where form appreciation can be seen in the divergence or widening of the channel (Hubel & Wiesel 1965). In the skin senses, a similar analysis has yet to be made empirically, although it has been subject to speculation.

The anterolateral-column pathways are a phylogenetically primitive system that receives impulses from unmyelinated and fine myelinated fibers; that conducts impulses which inform the brain of the state of the skin and tissues and, to some degree, of the ambient environment and objects (e.g., their warmth or coldness) but that may report poorly on spatial aspects of the stimulus. Geometrical patterns of warmed or cooled skin induced by radiant energy are very poorly discriminated. It may be important to know that a hand or an arm is cold or painful but not that one millimeter of skin is cold and the next warm. In contrast, the posterior-column pathways have been considered to convey knowledge of the objects forming the external world, the position of the body and the limbs in space, and the accurate locus of stimulation (Ruch & Patton 1965).

In this generalization, the reservation should be made that the anterolateral system of man and other primates is not the primitive system of fibers seen in lower forms, but it does reach the thalamus and conveys some limited knowledge of the world. Because the posterior columns are rarely selectively interrupted by trauma or tumor, the residual capacity of the spinothalamic and other lateral, anterolateral, and ventral ascending pathways is poorly understood. From experiments on monkeys (DeVito et al. 1964), it is known that the discrimination of weights is initially impaired by section of the posterior column but that the capacity to discriminate returns virtually to normal. Weight discrimination is, from clinical evidence, considered to be served primarily by posterior-column systems.

**The somatosensory cortex.** Buried in the central fissure which divides the parietal from the frontal lobe is a narrow strip of highly granular cortex typical of sensory areas in the cortex. On this and the adjacent zone of the postcentral gyrus is concentrated a dense projection of fibers from so-called relay nuclei, which in turn receive fibers of the medial lemniscus and the spinothalamic tract (in primates). This cortical area, called the somato-

sensory cortex, is highly organized topographically, with relatively large portions devoted to the skin areas having the greatest acuity in two-point discrimination and the greatest accuracy in location of a point touched. According to Mountcastle (1961) the receptive field of a single neuron in this region is relatively small but is calculated to be one hundred times that of a second-order neuron. The size of the receptive field is smallest on the fingers and largest on the back, being roughly proportional to the spatial acuity of these regions.

**Detour systems.** In addition to the classical somatosensory pathway, there are three additional pathways, termed *detour systems* (Ruch & Patton 1965) because they, in a sense, bypass the relay nuclei of the classical pathway. One is a projection to the precentral gyrus from a thalamic nucleus which receives input from the cerebellum and, hence, is potentially connected with the various spinocerebellar pathways.

The second is a projection of fibers from the posterior thalamic nuclei to the posterior parietal lobule, a so-called association area, lying behind the termination of the classical pathway. The posterior thalamic nuclei are believed to receive impulses from the classic relay nucleus (some of the most posterior ones receive impulses directly from the medial and spinal lemniscus, but their cortical projection is not known with certainty). At the time these systems were being traced anatomically, behavioral experiments showed that the posterior parietal area could account for discrimination of weights independently of the precentral gyrus (Ruch et al. 1937). Furthermore, the combined destruction of both areas (parietal lobectomy) greatly impaired, but did not abolish, weight discrimination, even in chimpanzees. This and other evidence (Kruger & Porter 1958) indicates that the classical motor area serves a discriminative somatosensory function.

The third set of detour pathways is nonlemniscal and nonspinothalamic in the strict sense but is nonetheless anterolateral in its spinal course. Ascending fibers terminate in or give off collaterals (some are no doubt collaterals of spinothalamic fibers) to the bulbar reticular formation and, at midbrain levels, to the central gray matter, the superior colliculus, and the posterior thalamic nuclei near the medial geniculate body (Mehler et al. 1960). The thalamic nuclei medial to the relay of spinothalamic and lemniscal fibers also receive a stream of the ascending fibers of anterolateral origin, as well as impulses from multisynaptic pathways deep in the reticular formation.



The pathway has been termed the paleospinothalamic tract, to distinguish it from the direct, or neospinothalamic, tract (Bishop 1962). There is considerable reason to believe that the diffuse, multisynaptic, unmyelinated, primitive system of fibers is involved in affect and pain experience.

**Posterolateral tracts.** Throughout this discussion, we have drawn the classical contrast between posterior-column and anterolateral-column pathways. It is now clear that a posterolateral tract of the spinal cord also sends impulses to the somatosensory cortex, with relays in the lateral cervical nucleus. This tract is a fast-fiber system, impulses reaching the cerebral cortex more quickly than those conducted in the posterior columns. Some fibers of this system have restricted receptive fields, while others respond broadly in respect to place and kind of cutaneous stimulation. This system has been studied mainly electrophysiologically, and its significance to sensation relative to the posterior-column and anterolateral-column systems has yet to be assessed.

**Mechanisms of cerebral sensory function.** Four aspects of sensation can be arranged in a presumed order of phylogenetic development: affect, quality, intensity, and localization. *Affect* appears to develop at subcortical levels of the nervous system, probably as low as the midbrain. The *quality* of stimuli can be correctly identified (e.g., touch, warmth) even after extensive cortical lesions, and recognition of quality is probably a thalamic, as well as a cortical, function. The discrimination of fine degrees of *intensity* is a function of the "motor" and "association" areas adjacent to the classic somatosensory area of the postcentral gyrus in the cerebral cortex. Obviously, these two terms do not convey their sensory function.

In the ascending pathways, distortion of a simple stimulus-intensity-frequency-of-discharge arrangement arises from the tendency of second-order neurons and cortical neurons to fire repetitively in response to a single afferent stimulation. Nevertheless, the single thalamic units being fired by limb displacement reflect faithfully, in spikes per second, the degree of joint rotation (Mountcastle et al. 1963). Interestingly, the quantitative relationship between stimulus and response is a power function rather than the logarithmic function postulated by the classical Weber-Fechner law. This finding agrees with and supports the conclusion reached by Stevens (1961) on the basis of psychophysical experiments involving magnitude-scaling procedures and cross-modality ratio matching; such methods are powerful tools applicable widely in the social sciences. [See **PSYCHOPHYSICS**;

**SCALING**; and the biographies of **FECHNER** and **WEBER**, ERNST HEINRICH.]

A broad relation exists between accuracy in localization and the amount of topographical organization in a system, the number of cortical units devoted to an area of skin, and the size of the receptive fields. However, the extent of the receptive fields and of the area of cerebral cortex activated by stimulation of a point on the skin are much too great to permit consideration of any true point-to-point isomorphism between the skin and the somatosensory cortex. Corroborative evidence is provided by the fact that the ramifications of single sensory neurons are so extensive that neurons interdigitate, and that size of the receptive field becomes progressively larger higher up in the nervous system. There are counterparts in the auditory system and in the visual system, where optical errors, light scatter, and small eye movements are also hazards to a conception of a point-to-point relationship between the external world and the visual cortex. It may be speculated that all these errors could have been minimized in the course of evolution. The occurrence of these dispersal processes must mean that they contribute to, rather than detract from, acuity of spatial discrimination. It is probable that in all three sensory systems the nervous system behaves, not in terms of a mosaic of all-or-nothing patches of cortical activation, but in modal points of maximum cortical activation (number of active neurons times frequency of their discharge).

The ability to discriminate modal points of excitability and such devices for sharpening these points as lateral inhibition and descending inhibitory influences on sensory system synapses provide a mechanism by which the fineness of "cortical grain" (i.e., the numbers of cortical neurons) extends discriminative ability beyond what could be achieved if sense organs were arranged in a mosaic, rather than an interdigitating, fashion and were equipped with private pathways to the cerebral cortex. An alternative to this statistical conception of spatial discrimination is a very successful "atomistic" analysis of the visual cortex by Hubel and Wiesel (1965), in which divergence in the system at each synaptic level endows it with increasingly subtle powers of spatial movement and form discrimination.

THEODORE C. RUCH

[Directly related are the entries **NERVOUS SYSTEM**, article on **STRUCTURE AND FUNCTION OF THE BRAIN**; **PAIN**; **SENSES**, article on **CENTRAL MECHANISMS**. Other relevant material may be found in **BODY IMAGE**; **HEARING**; **PERCEPTION**, article on **DEPTH PER-**

CEPTION; PSYCHOLOGY, *article on* PHYSIOLOGICAL PSYCHOLOGY; PSYCHOPHYSICS; SENSES; TASTE AND SMELL; VISION.]

## BIBLIOGRAPHY

- BISHOP, GEORGE H. 1962 Normal and Abnormal Sensory Patterns: Pain. I: Anatomical, Physiological and Psychological Factors in Sensation of Pain. Pages 95-133 in Robert Grenell (editor), *Neural Physiopathology*. New York: Harper.
- BORING, EDWIN G. 1942 *Sensation and Perception in the History of Experimental Psychology*. New York: Appleton.
- DEVITO, J.; RUCH, T. C.; and PATTON, H. D. 1964 Analysis of Residual Weight Discriminatory Ability and Evoked Cortical Potentials Following Section of Dorsal Columns in Monkeys. *Indian Journal of Physiology and Pharmacology* 8:117-126.
- DOUGLAS, W. W.; RITCHIE, J. M.; and STRAUB, R. W. 1960 The Role of Non-myelinated Fibres in Signalling Cooling of the Skin. *Journal of Physiology* 150:266-283.
- HEAD, HENRY et al. 1920 *Studies in Neurology*. 2 vols. London: Hodder & Stoughton. → These papers consist mainly of a republication of papers published in *Brain* between 1905 and 1918.
- HENSEL, HERBERT; and BOMAN, KURT K. A. 1960 Afferent Impulses in Cutaneous Sensory Nerves in Human Subjects. *Journal of Neurophysiology* 23:564-578.
- HUBEL, DAVID H.; and WIESEL, TORSTON N. 1965 Receptive Fields and Functional Architecture in Two Non-striate Visual Areas (18 and 19) of the Cat. *Journal of Neurophysiology* 28:229-289.
- IGGO, A. 1959 Cutaneous Heat and Cold Receptors With Slowly Conducting (C) Afferent Fibres. *Quarterly Journal of Experimental Physiology* 44:362-370.
- IGGO, A. 1965 The Peripheral Mechanisms of Cutaneous Sensation. Pages 92-100 in David R. Curtis and A. K. McIntyre (editors), *Studies in Physiology, Presented to John C. Eccles*. New York: Springer.
- IGGO, A.; and MUIR, A. R. 1963 A Cutaneous Sense Organ in the Hairy Skin of Cats. *Journal of Anatomy* 97:151 only. → An abstract.
- KRUGER, LAWRENCE; and PORTER, PHYLLIS 1958 A Behavioral Study of the Functions of the Rolandic Cortex in the Monkey. *Journal of Comparative Neurology* 109:439-469.
- MEHLER, WILLIAM R.; FEFFERMAN, MARTIN E.; and Nauta, WALLE J. H. 1960 Ascending Axon Degeneration Following Anterolateral Cordotomy: An Experimental Study in the Monkey. *Brain* 83:718-750.
- MOUNTCASTLE, VERNON B. 1961 Some Functional Properties of the Somatic Afferent System. Pages 403-436 in Symposium on Principles of Sensory Communication, Endicott House, 1959, *Sensory Communication*. Cambridge, Mass.: M.I.T. Press.
- MOUNTCASTLE, VERNON B.; POGGIO, GIAN F.; and WERNER, GERHARD 1963 The Relation of Thalamic Cell Response to Peripheral Stimuli Varied Over an Intensive Continuum. *Journal of Neurophysiology* 26:807-834.
- POGGIO, GIAN F.; and MOUNTCASTLE, VERNON B. 1960 A Study of the Functional Contributions of the Lemniscal and Spinothalamic Systems to Somatic Sensibility. Johns Hopkins Hospital, *Bulletin* 106:266-316.
- POWELL, THOMAS P.; and MOUNTCASTLE, VERNON B. 1959 Some Aspects of the Functional Organization of the Cortex of the Postcentral Gyrus of the Monkey: A Correlation of Findings Obtained in a Single Unit Analysis With Cytoarchitecture. Johns Hopkins Hospital, *Bulletin* 105:133-162.
- ROSE, JERZY E.; and MOUNTCASTLE, VERNON B. 1959 Touch and Kinesthesia. Volume 1, section 1, pages 387-429 in *Handbook of Physiology*. Washington: American Physiological Society.
- RUCH, THEODORE C.; FULTON, J. F.; and GERMAN, W. J. 1937 Sensory Discrimination in Monkey, Chimpanzee and Man After Lesions on the Parietal Lobe. *Archives of Neurology and Psychiatry* 39:919-937.
- RUCH, THEODORE C.; and PATTON, HARRY D. (editors) 1965 *Physiology and Biophysics*. 19th ed. Philadelphia: Saunders. → The first edition was published in 1905 as *A Textbook of Physiology for Medical Students and Physicians*, by William H. Howell.
- SHERINGTON, CHARLES S. (1906) 1948 *The Integrative Action of the Nervous System*. 2d ed. New Haven: Yale Univ. Press.
- STEVENS, S. S. 1961 The Psychophysics of Sensory Function. Pages 1-33 in Symposium on Principles of Sensory Communication, Endicott House, 1959, *Sensory Communication*. Edited by Walter A. Rosenblith. Cambridge, Mass.: M.I.T. Press.
- WALL, PATRICK D. 1961 Two Transmission Systems for Skin Sensations. Pages 475-496 in Symposium on Principles of Sensory Communication, Endicott House, 1959, *Sensory Communication*. Edited by Walter A. Rosenblith. Cambridge, Mass.: M.I.T. Press.
- WEDDELL, GRAHAM 1961 Receptors for Somatic Sensation. Volume 1, pages 13-48 in Conference on Brain and Behavior, *Brain and Behavior*. Edited by Mary A. B. Brazier. Washington: American Institute of Biological Sciences.

## SLAVERY

In classical Roman law, slavery was defined as an institution "whereby someone is subject to the dominium of another contrary to nature" (*Digest* 1.5.4.1). *Dominium* can be translated as "power," but the idea of property is also implied. This definition may be accepted as universally applicable without the controversial phrase, "contrary to nature." Distinctions then have to be drawn according to the owner (whether an individual, a corporate institution, the state, or a god), according to the existence or nonexistence of certain "rights" of the slave (such as the claim to eventual manumission, or statutory freedom), and according to the social structure within which slavery functioned. However, the property element remains essential. All forms of labor on behalf of another, whether "free" or "unfree," place the man who labors in the power of another; what separates the slave from the rest, including the serf or peon, is the totality of his powerlessness in principle, and for that the idea of property is juristically the key—hence the term "chattel slave."

For a sociological analysis, however, equal stress must be given to the slave's deracination. The law



The pathway has been termed the paleospinothalamic tract, to distinguish it from the direct, or neospinothalamic, tract (Bishop 1962). There is considerable reason to believe that the diffuse, multisynaptic, unmyelinated, primitive system of fibers is involved in affect and pain experience.

**Posterolateral tracts.** Throughout this discussion, we have drawn the classical contrast between posterior-column and anterolateral-column pathways. It is now clear that a posterolateral tract of the spinal cord also sends impulses to the somatosensory cortex, with relays in the lateral cervical nucleus. This tract is a fast-fiber system, impulses reaching the cerebral cortex more quickly than those conducted in the posterior columns. Some fibers of this system have restricted receptive fields, while others respond broadly in respect to place and kind of cutaneous stimulation. This system has been studied mainly electrophysiologically, and its significance to sensation relative to the posterior-column and anterolateral-column systems has yet to be assessed.

**Mechanisms of cerebral sensory function.** Four aspects of sensation can be arranged in a presumed order of phylogenetic development: affect, quality, intensity, and localization. *Affect* appears to develop at subcortical levels of the nervous system, probably as low as the midbrain. The *quality* of stimuli can be correctly identified (e.g., touch, warmth) even after extensive cortical lesions, and recognition of quality is probably a thalamic, as well as a cortical, function. The discrimination of fine degrees of *intensity* is a function of the "motor" and "association" areas adjacent to the classic somatosensory area of the postcentral gyrus in the cerebral cortex. Obviously, these two terms do not convey their sensory function.

In the ascending pathways, distortion of a simple stimulus-intensity-frequency-of-discharge arrangement arises from the tendency of second-order neurons and cortical neurons to fire repetitively in response to a single afferent stimulation. Nevertheless, the single thalamic units being fired by limb displacement reflect faithfully, in spikes per second, the degree of joint rotation (Mountcastle et al. 1963). Interestingly, the quantitative relationship between stimulus and response is a power function rather than the logarithmic function postulated by the classical Weber-Fechner law. This finding agrees with and supports the conclusion reached by Stevens (1961) on the basis of psychophysical experiments involving magnitude-scaling procedures and cross-modality ratio matching; such methods are powerful tools applicable widely in the social sciences. [See **PSYCHOPHYSICS**;

**SCALING**; and the *biographies* of FECHNER and WEBER, ERNST HEINRICH.]

A broad relation exists between accuracy in localization and the amount of topographical organization in a system, the number of cortical units devoted to an area of skin, and the size of the receptive fields. However, the extent of the receptive fields and of the area of cerebral cortex activated by stimulation of a point on the skin are much too great to permit consideration of any true point-to-point isomorphism between the skin and the somatosensory cortex. Corroborative evidence is provided by the fact that the ramifications of single sensory neurons are so extensive that neurons interdigitate, and that size of the receptive field becomes progressively larger higher up in the nervous system. There are counterparts in the auditory system and in the visual system, where optical errors, light scatter, and small eye movements are also hazards to a conception of a point-to-point relationship between the external world and the visual cortex. It may be speculated that all these errors could have been minimized in the course of evolution. The occurrence of these dispersal processes must mean that they contribute to, rather than detract from, acuity of spatial discrimination. It is probable that in all three sensory systems the nervous system behaves, not in terms of a mosaic of all-or-nothing patches of cortical activation, but in modal points of maximum cortical activation (number of active neurons times frequency of their discharge).

The ability to discriminate modal points of excitability and such devices for sharpening these points as lateral inhibition and descending inhibitory influences on sensory system synapses provide a mechanism by which the fineness of "cortical grain" (i.e., the numbers of cortical neurons) extends discriminative ability beyond what could be achieved if sense organs were arranged in a mosaic, rather than an interdigitating, fashion and were equipped with private pathways to the cerebral cortex. An alternative to this statistical conception of spatial discrimination is a very successful "atomistic" analysis of the visual cortex by Hubel and Wiesel (1965), in which divergence in the system at each synaptic level endows it with increasingly subtle powers of spatial movement and form discrimination.

THEODORE C. RUCH

[Directly related are the entries **NERVOUS SYSTEM**, article on **STRUCTURE AND FUNCTION OF THE BRAIN**; **PAIN**; **SENSES**, article on **CENTRAL MECHANISMS**. Other relevant material may be found in **BODY IMAGE**; **HEARING**; **PERCEPTION**, article on **DEPTH PER-**

CEPTION; PSYCHOLOGY, article on PHYSIOLOGICAL PSYCHOLOGY; PSYCHOPHYSICS; SENSES; TASTE AND SMELL; VISION.]

#### BIBLIOGRAPHY

- BISHOP, GEORGE H. 1962 Normal and Abnormal Sensory Patterns: Pain. I: Anatomical, Physiological and Psychological Factors in Sensation of Pain. Pages 95-133 in Robert Grenell (editor), *Neural Physiopathology*. New York: Harper.
- BORING, EDWIN G. 1942 *Sensation and Perception in the History of Experimental Psychology*. New York: Appleton.
- DEVITO, J.; RUCH, T. C.; and PATTON, H. D. 1964 Analysis of Residual Weight Discriminatory Ability and Evoked Cortical Potentials Following Section of Dorsal Columns in Monkeys. *Indian Journal of Physiology and Pharmacology* 8:117-126.
- DOUGLAS, W. W.; RITCHIE, J. M.; and STRAUB, R. W. 1960 The Role of Non-myelinated Fibres in Signalling Cooling of the Skin. *Journal of Physiology* 150:266-283.
- HEAD, HENRY et al. 1920 *Studies in Neurology*. 2 vols. London: Hodder & Stoughton. → These papers consist mainly of a republication of papers published in *Brain* between 1905 and 1918.
- HENSEL, HERBERT; and BOMAN, KURT K. A. 1960 Afferent Impulses in Cutaneous Sensory Nerves in Human Subjects. *Journal of Neurophysiology* 23:564-578.
- HUBEL, DAVID H.; and WIESEL, TORSTON N. 1965 Receptive Fields and Functional Architecture in Two Non-striate Visual Areas (18 and 19) of the Cat. *Journal of Neurophysiology* 28:229-289.
- ICCO, A. 1959 Cutaneous Heat and Cold Receptors With Slowly Conducting (C) Afferent Fibres. *Quarterly Journal of Experimental Physiology* 44:362-370.
- ICCO, A. 1965 The Peripheral Mechanisms of Cutaneous Sensation. Pages 92-100 in David R. Curtis and A. K. McIntyre (editors), *Studies in Physiology*. Presented to John C. Eccles. New York: Springer.
- ICCO, A.; and MUIR, A. R. 1963 A Cutaneous Sense Organ in the Hairy Skin of Cats. *Journal of Anatomy* 97:151 only. → An abstract.
- KRUGER, LAWRENCE; and PORTER, PHYLLIS 1958 A Behavioral Study of the Functions of the Rolandic Cortex in the Monkey. *Journal of Comparative Neurology* 109:439-469.
- MEHLER, WILLIAM R.; FEFERMAN, MARTIN E.; and Nauta, WALLE J. H. 1960 Ascending Axon Degeneration Following Anterolateral Cordotomy: An Experimental Study in the Monkey. *Brain* 83:718-750.
- MOUNTCASTLE, VERNON B. 1961 Some Functional Properties of the Somatic Afferent System. Pages 403-436 in Symposium on Principles of Sensory Communication, Endicott House, 1959, *Sensory Communication*. Cambridge, Mass.: M.I.T. Press.
- MOUNTCASTLE, VERNON B.; POGGIO, GIAN F.; and WERNER, GERHARD 1963 The Relation of Thalamic Cell Response to Peripheral Stimuli Varied Over an Intensive Continuum. *Journal of Neurophysiology* 26:807-834.
- POGGIO, GIAN F.; and MOUNTCASTLE, VERNON B. 1960 A Study of the Functional Contributions of the Lemniscal and Spinothalamic Systems to Somatic Sensibility. Johns Hopkins Hospital, *Bulletin* 106:266-316.
- POWELL, THOMAS P.; and MOUNTCASTLE, VERNON B. 1959 Some Aspects of the Functional Organization of the Cortex of the Postcentral Gyrus of the Monkey: A Correlation of Findings Obtained in a Single Unit Analysis With Cytoarchitecture. Johns Hopkins Hospital, *Bulletin* 105:133-162.
- ROSE, JERZY E.; and MOUNTCASTLE, VERNON B. 1959 Touch and Kinesthesia. Volume 1, section 1, pages 387-429 in *Handbook of Physiology*. Washington: American Physiological Society.
- RUCH, THEODORE C.; FULTON, J. F.; and GERMAN, W. J. 1937 Sensory Discrimination in Monkey, Chimpanzee and Man After Lesions on the Parietal Lobe. *Archives of Neurology and Psychiatry* 39:919-937.
- RUCH, THEODORE C.; and PATTON, HARRY D. (editors) 1965 *Physiology and Biophysics*. 19th ed. Philadelphia: Saunders. → The first edition was published in 1905 as *A Textbook of Physiology for Medical Students and Physicians*, by William H. Howell.
- SHERRINGTON, CHARLES S. (1906) 1948 *The Integrative Action of the Nervous System*. 2d ed. New Haven: Yale Univ. Press.
- STEVENS, S. S. 1961 The Psychophysics of Sensory Function. Pages 1-33 in Symposium on Principles of Sensory Communication, Endicott House, 1959, *Sensory Communication*. Edited by Walter A. Rosenblith. Cambridge, Mass.: M.I.T. Press.
- WALL, PATRICK D. 1961 Two Transmission Systems for Skin Sensations. Pages 475-496 in Symposium on Principles of Sensory Communication, Endicott House, 1959, *Sensory Communication*. Edited by Walter A. Rosenblith. Cambridge, Mass.: M.I.T. Press.
- WEDDELL, GRAHAM 1961 Receptors for Somatic Sensation. Volume 1, pages 13-48 in Conference on Brain and Behavior, *Brain and Behavior*. Edited by Mary A. B. Brazier. Washington: American Institute of Biological Sciences.

#### SLAVERY

In classical Roman law, slavery was defined as an institution "whereby someone is subject to the dominium of another contrary to nature" (*Digest* 1.5.4.1). *Dominium* can be translated as "power," but the idea of property is also implied. This definition may be accepted as universally applicable without the controversial phrase, "contrary to nature." Distinctions then have to be drawn according to the owner (whether an individual, a corporate institution, the state, or a god), according to the existence or nonexistence of certain "rights" of the slave (such as the claim to eventual manumission, or statutory freedom), and according to the social structure within which slavery functioned. However, the property element remains essential. All forms of labor on behalf of another, whether "free" or "unfree," place the man who labors in the power of another; what separates the slave from the rest, including the serf or peon, is the totality of his powerlessness in principle, and for that the idea of property is juristically the key—hence the term "chattel slave."

For a sociological analysis, however, equal stress must be given to the slave's deracination. The law



may declare him, in a formal way, powerless and rightless; one reason the law is enforceable is that he lacks any counterweight or support, whether from a religious institution, from a kinship group, from his own state or nation, or even from other depressed groups within the society in which he has become a slave. Legally he is not a person. Yet he is a human being, and therefore a purely juristic analysis in terms of property, though necessary, is not sufficient.

Conceptually, every man has available to him, or is denied, a bundle of rights and obligations as diverse as freedom of movement, the right to the fruits of his labor, the right to marry and establish a family, the obligation (or right) of military service, the right to look after his soul. It is not normally the case that a man possesses either all of them or none; hence the range and variety of personal statuses found in different societies, and, within limits, even inside a single society, are very considerable. One may speak of a spectrum of statuses between the two extremes of absolute rightlessness and of absolute freedom to exercise all rights at all times (Finley 1964). The latter has never existed, nor has the former, although the position of the slave in the American South came very near to it. In between the two extremes, precisely as in a spectrum, there is much shading and overlapping, which the servile vocabulary reflects.

Within the spectrum there are lines of demarcation. Throughout most of human history, labor for others has been performed in large part under conditions of dependence or bondage; that is to say, the relation between the man who works and his master or employer rested neither on ties of kinship nor on a voluntary, revocable contract of employment, but rather on birth into a class of dependents, on debt, or on some other precondition which by custom and law automatically removed from the dependent, usually for a long term or for life, some measure of his freedom of choice and action. "The concept of labor as a salable commodity, apart from the person of the seller, is relatively recent in the history of civilization" (Lasker 1950, p. 114). In all societies in which dependent labor is common, regardless of the variations within that broad class of persons, one main demarcation line is between the dependents and the others.

Slavery is a species of dependent labor and not the genus. Slaves were to be found in many societies in which other kinds of dependent labor—debt bondsmen, clients, helots, serfs, Babylonian *mushkenū*, Chinese *k'o*, Indian Sudras—were common, just as they coexisted with free labor. However, slavery attained its greatest functional significance,

and usually its greatest numerical strength, in societies in which other, less total varieties of bondage had either disappeared or had never existed. The distinction is particularly sharp as between *genuine* slave societies—classical Greece (except Sparta) and Rome, the American South and the Caribbean—on the one hand, and *slave-owning* societies as found in the ancient Near East (including Egypt), India, or China, on the other hand. Only when slaves became the main dependent labor force was the concept of personal freedom first articulated (in classical Greece), and words were then created or adapted to express that idea. It is literally impossible to translate the word "freedom" directly into ancient Babylonian or classical Chinese, and modern European languages cannot render *mushkenum* or *k'o*.

### The slave as outsider

Speculations about the origins of slavery have tended to overlook the specific character of slavery within the broader category of dependent labor. Thus, Nieboer (1900, especially pp. 6–7, 419–430) correctly stresses "division of labor" and scarce resources (including scarce manpower) as necessary conditions, but he misses the point that the needs were usually met in early and traditional societies by such institutions as clientage or helotage rather than by slavery, as they have again been met in many areas in very modern times after the abolition of slavery (Kloosterboer 1960, chapter 14).

The slave is an outsider: that alone permits not only his uprooting but also his reduction from a person to a thing which can be owned. Insiders en masse cannot be so totally transformed; no community could survive that. Thus, free Greeks who wished to dispose of unwanted children were compelled to resort to the fiction that they had "exposed" them (that is, abandoned them in a deserted place); the earliest Roman law code explicitly provided that if a Roman were subject to enslavement as a punishment, he had to be sold abroad (Lévy-Bruhl 1931); Islamic law always laid down, and usually enforced, the rule that no born Muslim could be enslaved.

Any hypothesis about the origins of slavery must therefore explain how and why a given society turned to outsiders either to supplement or to replace its existing labor force. Supplementation on a small scale, such as the retention of female captives, seems both very ancient and very widespread and presents no analytical problems. But the shift to slavery in a fundamental sense, as a substantial labor force employed in production, is a radical step. The explanations cannot be identical in all

instances, because of profound differences in the social structures and economic systems. However, there were always present not only a sufficient material and technical level and a concentration of power which made possible safe procurement of outsiders in sufficient numbers but also the failure, unacceptability, or unavailability of other kinds of labor.

The trauma of enslavement, often entailing great physical suffering as well as severe psychological damage, set up a chain reaction in the behavior of both the slaves and their masters, in which the potential or actual employment of naked force was a permanent and inescapable factor (Elkins 1959). These behavior patterns and their underlying psychology were reinforced by the slave's lack of essential human ties of kin and community. Free sexual access to slaves marks them off from all other persons as much as their juridical classification as property. On the other hand, not all societies went as far as the American South in the absolute denial to the slave of a *de facto* family of his own. There slavery was complete, so to speak, and the slave's deprivation was extended to the next generation; he lost all control not only over his productive activity but also over his reproduction. In consequence, being born into slavery meant being born an outsider, too.

**Prejudice.** Prejudices of color, race, nationality, and religion were deeply involved in slavery, not only as ideological justification but also as influences on its institutional development. "Slavery was not born of racism," writes Williams (1944, p. 7), "rather, racism was the consequence of slavery." However, the question must be asked whether the very idea of enslavement could have been thought of without the extreme distinction between groups, and therefore prejudice, in which "race" in a very loose sense was the criterion. To be sure, Greeks enslaved Greeks from other city-states, for example, and religious conversion, whether to Christianity or to Islam, did not normally release a slave. Nor did community solidarity always prevent penal bondage from sliding into genuine slavery (Pulleyblank 1958, p. 204). These are minor aberrations, however. If one could compile statistics of the number of slaves throughout history according to their origins, the proportion of racial, national, and religious outsiders would be overwhelming. Prejudice was certainly an important factor in the Southern American colonies when they decreed, in the 1660s, that henceforth all Negroes, but no whites, who were imported should be slaves and not indentured servants.

Prejudice had its limits, however. For example,

it never interfered with sexual relations. It allowed Portuguese officials and missionaries to condone Negro slavery in Brazil while they struggled energetically to emancipate the Amerindians (Boxer 1963, chapter 3; Davis 1966, chapter 8). Slaves drawn from culturally advanced peoples, such as the Hellenized Syrians in Rome, were regularly employed in such occupations as medicine and education. The most remarkable groups of elite slaves—the Mamelukes and Janissaries—illustrate all aspects of the slave outsider. In each generation the Mamelukes were purchased as children outside Islam, were given a rigorous and lengthy religious and military training, and were freed when ready for military service. A closed corps was thus created; their only ties were to themselves and their patron (ex-owner), and their elite position was not transmissible to their own children (Ayalon 1960).

### Slave supply

The procurement of a continuous and numerous supply of slaves depended above all on warfare. In early and simple societies, that usually meant raids by the slave-owning society on its source of supply. Even under more advanced conditions, when societies of more or less equal power and culture adjoined each other, regular warfare and raiding may also have been stimulated, at least in part, by the desire for slaves. However, greater stability of supply and greater numbers were ensured in the New World and even to a considerable extent in ancient and medieval times by a more indirect link with war. Neither Portugal nor England made war regularly in Africa in order to meet the demand in the Americas for slaves. The initial act of capture was left to the Africans themselves or to so-called pirates, as it had been left in antiquity to Scythians, Phrygians, and others. In short, the active cooperation of "native" chieftains and tribesmen was critical, and equally so was the role of professional slave traders as the middlemen.

Slave traders often appear as ambiguous figures. The Southern judge who wrote that "the calling of a slavetrader was always hateful, odious, even among slaveholders themselves" (Bancroft [1931] 1959, p. 366) was expressing one common judgment, but not the only one, for in England at the same time "his business was a recognized road to gentility" (Davis 1966, p. 154). In all countries his financial and governmental backers and his customers were thoroughly "respectable" figures in the community, and the high value of his services was always acknowledged. The suggestion that for a century or more the Roman Senate made no seri-



ous effort to suppress piracy in the eastern Mediterranean is probably sound, just as there can be little doubt about influential, though not unchallenged, support for the extensive illicit trade in slaves which followed British abolition after the Napoleonic Wars.

After warfare, breeding was the major source of supply. This is a subject on which much research remains to be done, the results of which will probably confirm the view that no simple generalization is possible. Certainly the often cited "law" that a slave population never reproduces itself is fictitious. In the United States the slaves did better than that, providing a very considerable increase. The question is intimately bound up with many social and economic factors and not with supposedly necessary demographic consequences (biological or otherwise) of the slave status. At one extreme there were conditions such as prevailed in the silver-mining district of Athens, where the slaves were almost all males and therefore could not reproduce themselves. At the other extreme there was the systematic, profitable breeding in the poorer regions of the American South (Conrad & Meyer 1958). In between these extremes, there was a great range of possibilities, conditioned by, among other things, the prevailing rules regarding the inheritance of slave status. These rules appear bewildering in their variety, but much the commonest was that the child took the mother's status.

### The uses of slaves

The actual numbers of slaves in any society are rarely known. The American South provides the decisive exception, and there the figures show an upper limit far below the often repeated exaggerations, such as the 400,000 claimed for ancient Athens. In 1860 slightly fewer than one-third of the population of the American slave states were slaves. Furthermore, "nearly three-fourths of all free Southerners had no connection with slavery through either family ties or direct ownership. The 'typical' Southerner was not only a small farmer but also a nonslaveholder" (Stampp 1956, p. 30). What counts in evaluating the place of slavery in any society is, therefore, not absolute totals or proportions, but rather *location* and *function*. If the economic and political elite depended primarily on slave labor for basic production, then one may speak of a slave society. It does not matter, in such situations, whether as many as three-fourths were not slaveholders, or whether slavery was fairly widespread outside the elite in domestic or other nonproductive roles.

Wherever there are slaves, they will be found in

domestic (and therefore also sexual) roles. Such roles have their own spectrum, ranging from the "drawers of water" and meanest prostitutes to domestics who were occupationally employed by their craftsmen-owners and to eunuchal grand viziers and harem favorites. If, however, this is the social location of most of the slaves, then it must follow that other kinds of dependent (or, on occasion, free) labor together with independent peasants and craftsmen constitute the productive labor force. That was the case in the ancient Near East, China, India, and medieval Europe and Byzantium as well as the Islamic world of the same period, and it is still the case in Saudi Arabia.

**The economics of slavery.** Slavery, then, is transformed as an institution when slaves play an essential role in the economy. Historically that has meant, in the first instance, their role in agriculture. Slavery has been accommodated to the large estate under radically different conditions: the Roman *latifundia* did not practice the monoculture of the modern plantation, and they existed within an essentially precapitalist economy. However, both types of estate produced for the market, and they both existed alongside widespread free small holding. That both slaves and free men did identical work was irrelevant; what mattered was the condition of the work, or rather, on whose behalf and under what (and whose) controls it was carried on. In slave societies hired labor was rare and slave labor the rule whenever an enterprise was too big for a family to conduct unaided. That rule extended from agriculture to manufacture and mining, and sometimes even to commerce and finance. In this article it is impossible to examine in detail these other uses of slaves, because of all the complexities involved and the extent to which they vary from society to society. A number of variables are involved: the poverty of the soil, as in Athens and other Greek cities; the special position of a particular region within an international network of economic relations, as in the American South; or the special role of the state as a large consumer of manufactures, as in the later Roman and Byzantine empires.

As an economic institution, slavery was "profitable"; this can be asserted with confidence, despite frequent attempts to deny it. In the strict sense of the term, the question of profitability does not enter into an evaluation of domestic slaves, court eunuchs and concubines, or Mamelukes. Nor is there any value in hypothetical arguments about whether or not Roman senators could have managed their *latifundia* even more profitably with some other kind of labor force. They made very large fortunes

for centuries on end, and there is no other way to calculate the economics of slavery in a precapitalist society. As for the American South, it can no longer be seriously questioned that slave plantations were profitable "in a strict accounting sense" (Genovese 1965, p. 280), whatever the effects of slavery on further economic growth within a competitive world economy. In the accounting, it is important to give proper weight to the profitability of slave breeding in the agriculturally poorer regions. In addition, there were the profits of the slave trade, which might or might not accrue to members of the slaveholding society itself.

### Slaves and masters

The difficulties in properly understanding the personality and the psychology of the slave are obvious. Neither the remarks by contemporary writers (whether slaveholders or outside reporters) nor the relatively few documents emanating from slaves themselves can be taken at face value. Yet a special slave psychology must have developed (speaking in group terms, of course). In order to survive as human beings, slaves had to adapt to their new state of deracination by developing new psychological features and new focuses of attachment, including their overseers and masters. Slave elites, whether individual overseers or whole groups of slaves and freedmen (ex-slaves), such as the imperial *familia* in Rome or the Mamelukes in Egypt, serve to exemplify how far adaptation and acceptance could be pushed under certain conditions. The slave-type—the clever schemer of Greek and Roman comedy or the childlike, indolent, amoral Sambo familiar to American literature and popular humor—is no doubt a stereotype and a caricature, but, as Elkins (1959, chapter 3) has argued, it cannot be a pure invention out of nothing.

**Slave rebellions.** The slave was a "troublesome property" (Stampp 1956, chapter 3). In its most extreme form, "being troublesome" meant revolt, but large-scale revolt is extremely difficult to organize and has, in fact, been a relatively rare phenomenon in the history of slavery. Throughout classical antiquity there were only three revolts of any mark, each involving 100,000 or more slaves, and all concentrated within the short time span of 135–70 B.C. Common to all three were the presence of certain necessary conditions, including a severe breakdown of the social order and the concentration of large numbers of slaves with common nationality, language, and culture, among them men with unusual potentialities of leadership (Vogt 1957). It is important to contrast the ancient chattel slaves with the helots (in Sparta and elsewhere)

in this respect: the latter were permanently mutilous in an organized way, presumably because they belonged to a class of dependent labor which retained the normal human ties of solidarity with kin and community. The Caribbean throughout the eighteenth and early nineteenth centuries was also an area of persistent revolt. In the United States, under conditions which differed above all by being noncolonial, not a single serious revolt ever occurred; for example, the "Turner cataclysm" of 1831 was a purely local affair lasting a few months from its inception (with only three days of actual fighting) and involving only some hundreds of men (Aptheker 1943, chapter 12).

"Being troublesome," in sum, usually meant something much less than outright rebellion, such as flight, sabotage, theft, and inefficiency. None of these is expressible in quantitative terms or easy to evaluate. There is American evidence to support the famous judgment of the economist Cairnes (1862) that slave labor was on the lowest level of skill because slaves were both uneducated and uncooperative (Genovese 1965, chapter 2). On the other hand, the possibilities of "loyalty," which is equally immeasurable, cannot be ignored. In contrast to American slaves, the slaves of ancient Greece and Rome were regularly and successfully employed in the most highly skilled occupations. Relative mildness or harshness of treatment cannot be a sufficient explanation of such variations, which must lie deep in the social structure and in psychology. Likewise, variations in the practice of manumission, in the place of freedmen in the society, and in the accompanying psychology require complex explanations.

**Attitudes of the masters.** In the ancient world the institution of slavery was never challenged, despite the notion that it was "contrary to nature." No serious argument was ever put forward for the abolition of slavery in ancient Greece and Rome (as distinct from relative liberality in freeing individual slaves), on moral or any other grounds; this was also the case in India, China, and the Islamic world. Nor did Christianity change the fundamental attitude after it became the official and more or less universal religion of both the western and eastern halves of the Roman Empire. Slavery declined sharply at the end of antiquity, but for reasons having nothing to do with moral ideas. Furthermore, it was in the Christian states in southern and southwestern Europe that slavery was considerably revived in the late Middle Ages (Verlinden 1955), and it was among the Christian conquerors of the New World that it received its newest and most vigorous re-creation. Paradoxically, it was



then that the most powerful and persistent claims were put forward for the "naturalness" of slavery, with ample quotation from the Bible, and that moral arguments for the *abolition* of slavery were fully mustered for the first time.

The whole subject of the psychological effects of slavery calls urgently for further investigation—from the side of the masters (including the free poor who themselves owned no slaves) as well as from the side of the slaves. The need to be brutal, ideologically as well as physically, must have had repercussions on the master's psyche. The easy sexual access to slave women influenced all attitudes toward sex and women: witness the quasi-chivalric ideology of Southern womanhood. Furthermore, the identification of certain forms of physical labor with slavery, including the essential labor in agriculture, had its effects on the free man's choice of employment and on his spirit of enterprise generally. More often than not, the majority of free farmers and craftsmen, out of necessity, performed labor similar to that of the slave. Even then, however, there were subtle effects on the directions into which creative talents and energies were channeled, and there were certain employments into which it was extremely difficult to move the free poor when they were needed. Policy makers in underdeveloped countries are still coming up against just such resistance (McLoughlin 1962), although it usually follows the abolition of forms of dependent labor other than slavery.

### Slavery and Marxist theory

Marxist theory, by its very nature, has assigned a unique historical position to slavery. History is viewed as a progress through a number of stages, each genetically determined within its predecessor and each founded on a particular mode of production (social relations of production), of which one is slavery. In the past half century, in particular, the way in which historical analysis was enmeshed in, not to say dominated by, current political discussions produced among orthodox Marxists a rigid, universal, unilinear scheme of development in five stages: primitive communism, slavery, feudalism, capitalism, and socialism. Even the ancient Near East and ancient China, it was held, were slave societies, and there were persistent but wholly unsuccessful efforts to discover general laws or general features common to all slave societies.

However, scattered through Marx's writings down to the early 1860s there are also brief, not fully developed references to an "Asiatic mode of production." His only attempt to examine this systematically was in one section of a bulky manuscript written in 1857–1858 but not published

until 1939–1941 and not widely known before the 1950s. In this sophisticated account, the Asiatic mode of production is characterized as one in which there was no private property in the land and in which a despotic government ruled over the village communities, whose members were in a condition of "general slavery" and who were therefore not slaves in the chattel sense at all. Publication of this work has sparked a very intense new discussion, following a hiatus of nearly a generation (Pećirka 1964). The discussion is still in an early and fluid state, but the general trend seems clear. It is argued that the stages of evolution in European history from which the traditional scheme was constructed do not constitute a model for world history at all but were, on the contrary, a unique development. As a corollary, the "Asiatic" mode of production has been found on other continents as well, for example in Bronze Age Greece and among the Incas. The place of slavery in Marxist theory thus seems to be undergoing a redefinition to fit a multilinear pattern of development.

M. I. FINLEY

[See also ASIAN SOCIETY, article on SOUTHEAST ASIA; INTERNMENT AND CUSTODY.]

### BIBLIOGRAPHY

No serious full-scale history of slavery exists in one book. The books and articles listed here, with the bibliographies they include, cover the field quite thoroughly. For classical antiquity, Wallon 1847 is still valuable for its rich documentation. The one modern book on antiquity, Westermann 1955, is recognized to be unsatisfactory; a better, though not systematic, introduction will be found in the 11 articles collected in Finley 1960. For the fullest survey of modern views since the work of David Hume, with a Marxist critique, see Lentsman 1963, Part 1. On the historiography of American slavery, see Elkins 1959, Chapter 1, and the massive literature cited in the notes in Davis 1966; for the new discussion initiated by Elkins, see Sio 1965. On the unresolved controversy about Latin American, and especially Brazilian, slavery, see Elkins 1959 for the view that there has been considerable amelioration, as contrasted with North American slavery, and Davis 1966, Chapter 8, for the opposing view. Summaries of current Marxist discussions of the "Asiatic mode of production," with particular reference to slavery, will be found in Pećirka 1964 and Chesneau 1965, both of whom write as Marxists. An important stimulant of this discussion has been the publication of Marx 1857–1858.

- APTHECKER, HERBERT 1943 *American Negro Slave Revolts*. New York: Columbia Univ. Press.
- AYALON, DAVID 1960 *Studies in Al-Jabarti. Part I: Notes on the Transformation of Mamluk Society in Egypt Under the Ottomans*. *Journal of the Economic and Social History of the Orient* 3:148–174, 275–325. → Provides references to Ayalon's other articles on the Mamelukes.
- BANCROFT, FREDERIC (1931) 1959 *Slave Trading in the Old South*. New York: Ungar.
- BOXER, CHARLES R. 1963 *Race Relations in the Portuguese Colonial Empire: 1415–1825*. Oxford: Clarendon.

- BROWNING, R. 1958 *Rabstvo v vizantiiskol imperii: 600–1200* (Slavery in the Byzantine Empire: 600–1200). *Vizantiiskii vremennik* 14:38–55.
- BUCKLAND, WILLIAM W. 1908 *The Roman Law of Slavery: The Condition of the Slave in Private Law From Augustus to Justinian*. Cambridge Univ. Press.
- CAIRNES, JOHN E. (1862) 1863 *The Slave Power: Its Character, Career, and Probable Designs; Being an Attempt to Explain the Real Issues Involved in the American Contest*. 3d ed. New York: Carleton.
- CHANANA, DEV RAJ (1960) 1961 *Slavery in Ancient India, as Depicted in Pali and Sanskrit Texts*. London: Collet.
- CHESNEAUX, JEAN 1965 *Où en est la discussion sur le mode de production asiatique?* *Pensée* 122:40–59.
- CONRAD, ALFRED H.; and MEYER, JOHN R. 1958 *The Economics of Slavery in the Ante Bellum South*. *Journal of Political Economy* 66:95–130, 442–443.
- DAVIS, DAVID BRION 1966 *The Problem of Slavery in Western Culture*. Ithaca, N.Y.: Cornell Univ. Press.
- ELKINS, STANLEY M. (1959) 1963 *Slavery: A Problem in American Institutional and Intellectual Life*. New York: Universal Library.
- FILLER, LOUIS (1960) 1963 *The Crusade Against Slavery: 1830–1860*. New York: Harper.
- FINLEY, M. I. (editor) 1960 *Slavery in Classical Antiquity: Views and Controversies*. Cambridge: Heffer.
- FINLEY, M. I. 1964 *Between Slavery and Freedom. Comparative Studies in Society and History* 6:233–249.
- GENOVESE, EUGENE D. 1965 *The Political Economy of Slavery: Studies in the Economy and Society of the Slave South*. New York: Pantheon.
- HADJINICOLAOU-MARAVA, ANNE 1950 *Recherches sur la vie des esclaves dans le monde byzantin*. Athens: L'Institut Français.
- HURD, JOHN C. 1858–1862 *The Law of Freedom and Bondage in the United States*. 2 vols. Boston: Little.
- IANNI, OCTAVIO 1962 *As metamorfoses do escravo: Apogeu e crise da escravatura no Brasil meridional*. São Paulo (Brazil): Difusão Européia do Livro.
- JAMES, CYRIL L. R. (1938) 1963 *The Black Jacobins: Toussaint L'Ouverture and the Santo Domingo Revolution*. 2d ed. rev. New York: Random House.
- KLOOSTERBOER, WILLEMINA 1960 *Involuntary Labour Since the Abolition of Slavery: A Survey of Compulsory Labour Throughout the World*. Leiden (Netherlands): Brill.
- LASKER, BRUNO 1950 *Human Bondage in Southeast Asia*. Chapel Hill: Univ. of North Carolina Press.
- LENTSMAN, IAKOV A. 1963 *Rabstvo v mikenskoi i gomeroskoi Gretsii* (Slavery in Mycenaean and Homeric Greece). Moscow: Akademiia Nauk SSSR. → Translated into German by Maria Bräuer-Pospelova as Ja. A. Lencman, *Die Sklaverei im mykenischen und homerischen Griechenland*, and published in 1966 by Steiner.
- LÉVY-BRUHL, H. (1931) 1960 *Théorie de l'esclavage*. Pages 151–169 in M. I. Finley (editor), *Slavery in Classical Antiquity: Views and Controversies*. Cambridge: Heffer. → First published as "Esquisse d'une théorie sociologique de l'esclavage à Rome."
- McLOUGHLIN, PETER F. M. 1962 *Economic Development and the Heritage of Slavery in the Sudan Republic*. *Africa* 32:355–391.
- MARX, KARL (1857–1858) 1965 *Pre-capitalist Economic Formations*. Edited by E. J. Hobsbawm. New York: International Publishers. → A partial translation of Marx's posthumously published *Grundrisse der Kritik der politischen Ökonomie*.
- MENDELSSOHN, ISAAC 1949 *Slavery in the Ancient Near East: A Comparative Study of Slavery in Babylonia, Assyria, Syria and Palestine; From the Middle of the Third Millennium to the End of the First Millennium*. Oxford Univ. Press.
- NIEBOER, HERMAN J. (1900) 1910 *Slavery as an Industrial System: Ethnological Researches*. The Hague: Nijhoff.
- PRČIRKA, J. 1964 *Die sowjetischen Diskussionen über die asiatische Produktionsweise und über die Sklavenerhalterformation*. *Eirene: Studia graeca et latina* (Prague) 3:147–169.
- POKORA, TIMOTEUS 1963 *Existierte in China eine Sklavenerhaltergesellschaft?* *Archiv orientální* 31:353–363.
- PULLEYBLANK, E. G. 1958 *The Origins and Nature of Chattel Slavery in China*. *Journal of the Economic and Social History of the Orient* 1:185–220.
- SIEGEL, BERNARD J. 1945 *Some Methodological Considerations for a Comparative Study of Slavery*. *American Anthropologist New Series* 47:357–392.
- SIO, ARNOLD A. 1965 *Interpretations of Slavery: The Slave Status in the Americas*. *Comparative Studies in Society and History* 7:289–308.
- STAMPP, KENNETH M. 1956 *The Peculiar Institution: Slavery in the Ante-bellum South*. New York: Knopf.
- URBACH, EFRAIM E. (1960) 1964 *The Laws Regarding Slavery as a Source for Social History of the Period of the Second Temple, the Mishnah and Talmud*. Volume 1, pages 1–94 in London, Institute of Jewish Studies, *Papers of the Institute* . . . Edited by J. G. Weiss. Jerusalem: Hebrew Univ., Magnes Press; distributed by the Oxford Univ. Press. → First published in Hebrew in the journal *Zion*.
- VERLINDEN, CHARLES 1955 *L'esclavage dans l'Europe médiévale. I: Péninsule ibérique–France*. Bruges (Belgium): Tempel.
- VOGT, J. (1957) 1965 *Struktur der antiken Sklavenerkriege*. Pages 20–60 in J. Vögt, *Sklaverei und Humanität*. Historia, Einzelschriften, Heft 8. Wiesbaden (Germany): Steiner.
- WADE, RICHARD C. 1964 *Slavery in the Cities: The South 1820–1860*. Oxford Univ. Press.
- WALLON, HENRI (1847) 1879 *Histoire de l'esclavage dans l'antiquité*. 2d ed. 3 vols. Paris: Hachette.
- WESTERMANN, WILLIAM L. 1955 *The Slave Systems of Greek and Roman Antiquity*. Philadelphia: American Philosophical Society.
- WILBUR, CLARENCE M. 1943 *Slavery in China During the Former Han Dynasty: 206 B.C.–A.D. 25*. Field Museum of Natural History, Anthropological Series, Vol. 34. Chicago: The Museum.
- WILLIAMS, ERIC (1944) 1961 *Capitalism and Slavery*. New York: Russell & Russell.

## SLEEP

Historically, there has been concern with two major questions about sleep: What specific mechanisms start, maintain, and terminate sleep? What functions does sleep serve? In addition, there has been considerable interest in the relationship of sleep to learning and performance, as well as in abnormalities of sleep behavior.



**Causes and processes.** The oldest theories about the causes of sleep postulated either congestion or anemia of the brain. More recent hypotheses have attributed sleep to arterial anoxemia, endocrine periodicities, functional neural blocks, or the accumulation of fatigue products. Piéron's idea of a "hypnotoxin" that accumulates during waking and is metabolized away during sleep (1913, p. 520) is one of the best-known explanations. A large number of biochemical studies on blood and other body fluids (reviewed by Kleitman 1939) have failed to turn up evidence in support of any of these hypotheses. The anoxia theory was controverted by the finding that the oxygen content of blood was normal during sleep and that cerebral blood flow was actually increased. The toxin theory was seriously challenged by the finding that Siamese twins with a common blood supply failed to show synchronized sleep-wake rhythms. However, knowledge of the biochemical processes taking place within the sleeping brain is not sufficient to accept or reject the causal influence of chemical factors. At present, these and similar theories of the causes of sleep are not in fashion.

**Electrophysiological mechanisms.** Since the 1930s, the development of electrophysiological techniques has facilitated new approaches to the processes of sleep. Interest has focused on the elucidation of the specific brain mechanisms responsible for the onset and termination of sleep. It is now well established that wakefulness results from activity in the ascending reticular system of the brain stem. However, the central mechanisms for the onset and maintenance of sleep are not so well understood.

The reticular activating system, which originates in the reticular formation of the lower brain stem and extends upward to the hypothalamus, subthalamus, thalamus, and cortex, receives collaterals from all sensory pathways as well as from the cortex. Stimulation of the reticular system by any of these sources or by such biological products as norepinephrine is accompanied by behavioral activation and alertness. Arousal is usually sustained after removal of the stimulus. Lack of stimuli or lack of stimulus variation has soporific effects. [See NERVOUS SYSTEM, article on STRUCTURE AND FUNCTION OF THE BRAIN.]

Since the publication of Kleitman's evolutionary theory of sleep, most physiologists have accepted the "stimulus deficiency" explanation of sleep (1939). The onset and maintenance of sleep are said to be a result of reduced afferent stimulation that deprives the cortex of its sensory raw material. The resulting "deactivation" of the reticular system is, indeed, an important sleep-inducing

mechanism. Everyday experience indicates that in man and animals the preparations for sleep include the suppression of external stimuli. Destruction of the reticular system produces sleep as well as electroencephalographic (EEG) patterns indicative of sleep.

More recent evidence, however, does not fully support this view of sleep as a passive phenomenon. In fact, it strongly suggests the existence of active, sleep-producing mechanisms in the central nervous system. Low-frequency stimulation of several brain locations from the lower brain stem to the thalamus produces sleep, in contrast to the arousing effects caused by high-frequency electrical stimulation of the reticular system. Barbiturate infusion at most brain levels induces sleep, but at some locations induces waking. It is true that changes in the internal milieu, such as reduced blood pressure and low body temperature, contribute to sleep and that familiar or unchanging external stimuli favor sleep, but these conditions are not sufficient to maintain sleep. There is behavioral and physiological evidence to support the hypothesis that there are two mechanisms in the brain for the control of sleep and waking: an activating and a deactivating system. The many degrees of wakefulness and sleep probably result from complex interactions between these two systems, involving the interplay of numerous structures in the nervous system.

**The sleep-wake rhythms.** Most living things show alternating activity and quiescence in response to the alternation of day and night. Higher animals show polyphasic or monophasic and diurnal or nocturnal patterns, depending on their adaptive requirements. These "circadian" (*circa, dies*) rhythms seem to be determined primarily by internal mechanisms but usually can be "reset" by environmental pacemakers, such as night and day. Some biological cycles are remarkably persistent, whereas others are easily modified by environmental changes.

The cerebral cortex may have a unique role in the maintenance of sustained sleep and waking. All animals capable of adjusting their sleep-wake cycles to the monophasic diurnal rhythm of man have a well-developed cortex. Removal of the cortex in such animals causes reversion to polyphasic sleep cycles.

The sleep-wake cycles of rats, rabbits, puppies, and human infants are polyphasic. Monkeys and higher mammals, as well as many birds, have monophasic diurnal cycles. Attempts to modify the diurnal rhythm in man include two types of manipulation: alteration of the phase and alteration of the period of the sleep-wake cycle. The phases of sleep

and waking can be easily modified (as in shifts to night work), but attempts to alter the 24-hour period to days of 12, 48, 21, or 28 hours have had only limited success. Some subjects adjusted to 21-hour or 28-hour days, but none adjusted to 12-hour or 48-hour cycles. A practical requirement for readjusting the phases of sleep and waking occurs when travelers move across lines of longitude. World travelers (especially on long-distance flights) experience phase-shift asynchrony in the day-night cycle, which may cause emotional distress and impaired performance. It seems that about two weeks are needed for a complete diurnal readjustment of physiological rhythms.

A number of physiological rhythms vary with man's sleep-wake cycle. Among these are heart rate, skin resistance, and muscle tonus. One measure of considerable interest is the diurnal temperature cycle. Temperature is normally maximal by day and at its lowest in the early hours of the morning. Several studies have shown that performance efficiency, especially speed, is highly correlated with the temperature cycle. The diurnal variation of both temperature and efficiency are probably related to the establishment of a circadian rhythm of excitement in the reticular system and hypothalamus. Prolonged disregard of normal biological rhythms may lead to genuine, although poorly understood, "rhythm diseases" anchored in disturbed autonomic functioning. [See TIME, article on PSYCHOLOGICAL ASPECTS.]

Sleep seems to be a parasympathetic phenomenon. During normal sleep, general metabolism slows down. Heat production and body temperature decrease, heart rate and respiration decrease, blood pressure is lowered, and  $\text{CO}_2$  tension increases. Kleitman (1939) has pointed out, however, that most of these changes could be simply a consequence of the prone position or of muscle relaxation, independent of sleep.

*Physiological cycling during sleep.* During the waking state, the resting normal human with eyes closed usually displays a continuous EEG pattern of 8 to 12 cycles per second, which is known as the "alpha" rhythm. As he drifts into sleep, the alpha rhythm disappears and is replaced by a low-voltage pattern with irregular frequency. During this phase, thresholds for responding to environmental stimuli are about the same as those in the waking state. As sleep continues, the background EEG voltage becomes higher and the frequency of the rhythms decreases, until the deepest stages of sleep are reached. In the deepest stages, the EEG record is composed of low-frequency, high-voltage "delta" rhythms, and it is very difficult to waken the subject. Approximately 90 minutes after the onset of

sleep, the high-voltage patterns disappear and are replaced by a low-voltage irregular phase similar to that seen at the beginning of sleep. This cycle of alternating slow high-voltage and faster low-voltage rhythms recurs with a period of about 90 minutes. The emergent low-voltage pattern has commonly been believed to be a "light" phase of sleep, but recent experiments show that thresholds for awakening are often very high during this stage. [See NERVOUS SYSTEM, article on ELECTROENCEPHALOGRAPHY.]

Other physiological measures also suggest a 90-minute period as a unit of sleep time. Changes in such physiological and behavioral variables as skin conductivity, heart rate, breathing rhythms, eye movements, and reports of dreaming occur cyclically and in phase with the EEG cycle. Increased heart rate, rapid eye movements, and reports of dreaming occur during the low-voltage phase.

The EEG of cats, rats, and monkeys also show alternating slow high-voltage and fast low-voltage patterns during sleep. Jouvet (1961) has shown that the high-voltage phase in cats is controlled by cortical mechanisms, and the low-voltage phase by pontine and limbic structures. EEG recordings obtained simultaneously from deep brain structures, cortex, and scalp suggest that in the cat the low-voltage phase of sleep involves hindbrain "sleep" and forebrain "wakefulness." The stage is accompanied by inhibition of motor activity, reduced heart rate and blood pressure, very high thresholds for awakening, and rapid eye movements. There is evidence, then, that during the low-voltage "dreaming" stage of sleep at least part of the brain is highly activated. But the very high response thresholds to auditory and other stimuli are difficult to interpret. Some studies show that auditory stimuli that warn of immediate and severe consequences are responded to in this stage as readily as in the other low-voltage phases of sleep. Thus, during dreaming the subject appears to be controlled by interoceptive stimuli and seems to block out external signals unless they have poignant significance.

The cyclical sequence of EEG patterns of sleep in man appears to be quite stable, both from night to night and from subject to subject. Disturbance of these patterns by "deprivation" of either high-voltage or low-voltage EEG phases causes increasing amounts of the disturbed phase to appear on succeeding nights, as if compensation were necessary. Continued deprivation of the low-voltage "dreaming" phase causes irritability and unease.

*Sleep loss and biological functions.* Is sleep a vital function? One way to test this is to deprive an organism of sleep. Several animal studies have shown that prolonged loss of sleep results in death.



A number of human subjects have stayed awake 200 hours or more without sustaining serious physical damage, but after 120 hours without sleep, most subjects developed reversible delirious or psychotic states characterized by visual hallucinations, delusions of persecution, disorientation, and confusion. The psychosis of sleep deprivation, which disappears after normal sleep, is a model mental illness that has considerable significance for experimental psychiatry.

If sleep plays a vital role in biological survival and psychological health, precisely what is this role? It has been suggested that sleep has three main functions: (a) restoring metabolic balance, (b) permitting recovery from muscular fatigue, (c) aiding neural reorganization. With few exceptions, studies of sleep loss have not given positive support to any of these hypotheses. Most systems of the body hold up remarkably well with loss of sleep.

*Metabolic effects.* One recent study of sleep loss, however (Luby et al. 1960), has shown striking metabolic effects. After the fourth day of sleep deprivation there was evidence of a drastic breakdown in the manufacture of adenosine triphosphate (ATP), the substance that provides energy for a host of biological processes. Thus, the high-energy transfer systems of cell metabolism showed gross impairment. This metabolic dysfunction was accompanied by rapid deterioration in personality and performance. These findings are congruent with the hypothesis that the catabolic processes that predominate during the waking state are compensated for during sleep; that sleep permits the restoration of energy and anabolic compensation.

*Effect on performance.* During sleep loss the subject shows deficits in performance that are similar to those observed with fatigue. The deficit takes the form of lapses (brief periods of poor performance), which in the sleep-deprived subject are accompanied by extreme drowsiness, intrusive thoughts, visual illusions, and dreams. They are preceded and accompanied by a slowing of heart rate, breathing, and EEG rhythms. Lapses increase in frequency, duration, and depth as the period without sleep progresses until the subject reaches a kind of hypnagogic state in which the background EEG pattern resembles a light phase of sleep. The frequency of lapses during sleep loss is increased by monotony or by prolonged performance and is decreased by incentive. The main effects of these brief periods of extreme drowsiness are slowed reactions. The subject is usually able to sustain accuracy if given sufficient time. Tasks extremely sensitive to sleep loss are those that are prolonged,

repetitive, and work-paced, impose a high speed-load, and provide low incentive (Williams et al. 1959).

As sleep loss increases, other mental functions are affected. For example, the ability to store and retrieve new information is greatly impaired. Finally, when most data-processing functions are seriously debased, the subject becomes psychotic. Some goal-directed, sequential, and rational behavior remains, however. Even after seven days of sleep deprivation, subjects can pull themselves together for short-term problem-solving and can carry out well-learned, perhaps automatic, sensory-motor tasks.

Performance decrement and personality aberrations during sleep loss follow the diurnal temperature cycle. Effectiveness reaches a low point late at night (when body temperature is at its lowest) and improves during the next day, even though the amount of sleep loss has increased.

*Interaction with other stress.* There are at present very few reports on the combined effects of sleep loss and other types of stress. The available studies indicate that the effects are not readily predictable from a knowledge of the independent effects of each form of stress. For example, Wilkinson (1963) has shown that high-intensity background noise normally causes impaired performance on vigilance and choice reaction-time tasks. But the degrading effect of noise is reduced after a night of sleep loss. Apparently, noise and sleep deprivation produce different types of fatigue. With sleep loss, the level of stimulation required for arousal is too low; with noise, it may be too high. Similar complexities appear in studies that combine raised body temperature with sleep loss. Recent unpublished studies suggest that for the sleep-deprived subject a moderate increase in body temperature (say, one degree) improves performance, but that higher temperatures act synergistically with sleep deprivation to cause severe decrement in both speed and accuracy.

*Fatigue at skilled tasks.* The word "fatigue" usually refers to the changes in muscle metabolism that occur with prolonged exertion. A vast amount of research has been done on this topic, but it will not be discussed in this article. There are, however, many prolonged tasks requiring very little muscular effort, during which performance declines and the subjects report fatigue. The general nature of impairment at such tasks was first described by Bills (1931). He observed that performance did not decline as a simple function of time at the task, but instead became increasingly uneven. Adequate performance was interrupted from time to

time by "blocks," or brief periods of no response, which increased in frequency and duration with continued mental work. Bills attributed mental blocks to "a recurrent low condition of mental functioning." [See *FATIGUE*.]

In recent years the prototype task for the study of prolonged performance has been the so-called vigilance test. During such a task, an observer watches or listens for long periods for critical signals that occur from time to time in a background of neutral signals. Radar and sonar operations are practical examples of such tasks. During the course of an hour's performance, the proportion of signals missed is low at first, but after 15 to 30 minutes increases rapidly. These errors of omission are increased by monotony, by distraction, and by states of physiological depression such as those caused by sleepiness, low oxygen pressure, high nitrogen pressure, and tranquilizing drugs. Performance improves under incentive, and the drug amphetamine largely prevents the decline in efficiency. Thus the effects of prolonged performance are similar to those seen with sleep loss. Presumably, the decrement is due to recurrent periods of lowered cerebral vigilance. It is not certain whether these phases represent increased distractibility or simply drowsiness, or both. Very few studies have used continuous physiological monitoring during prolonged performance, but the available reports suggest that the physiological changes accompanying performance decrement are similar to those that accompany the lapses of sleep deprivation. [See *ATTENTION*.]

**Learning and performance during sleep.** There is no convincing evidence that learning can occur during sleep. The sleeping state seems to be incompatible with many forms of cognitive behavior. Simon and Emmons (1955) in a well-controlled series of experiments showed by continuous EEG monitoring that humans failed to learn items of information during actual sleep. Their results challenge all earlier studies purporting to show that learning can be induced during sleep, and they cast great doubt on the claims of numerous commercial enterprises that advertise sleep-learning methods.

However, negative results in one well-controlled study do not prove that all forms of learning are impossible in the sleeping state. Failure to induce learning may have been due to procedural problems. Simon and Emmons used only one night of training. There was no effort to "shape" complex responses from simpler behavior or to elucidate the precise stimulus attributes, incentive conditions, and response requirements that might be compatible with the sleeping state. The cyclic rise and fall

of EEG activation patterns during sleep suggests that there may be some phases of sleep in which learning is possible.

The ancient observation that some discrimination is possible in sleep has been confirmed by recent experiments. Cats are capable of remarkably fine discrimination of auditory signals while asleep. Humans can discriminate between complex auditory patterns, make appropriate motor and physiological responses to conditioned stimuli, respond differentially to warning signals and neutral signals, and perform well-learned motor sequences without awakening.

**Sleep disorders and sleep therapy.** The many abnormalities of sleep are well known. They include excessive sleep, inability to sleep, restless sleep, nightmares, bed-wetting, sleep paralysis, and other problems. Neurologists classify sleep disorders under three headings: hypersomnia, insomnia, and nocturnal behavioral symptoms. All three can occur with brain disease or with psychological etiology.

**Hypersomnia.** Narcolepsy is a condition of hypersomnia characterized by sudden sleep seizures at inappropriate times. The syndrome also includes sleep paralysis and cataplexy. The irresistible desire to sleep may occur several times daily, especially after heavy meals, during periods of low body temperature, or during monotonous activity. The duration of sleep varies from a few seconds to several hours. Hypnagogic hallucinations often precede and accompany the attacks of sleep.

Narcolepsy occurs in both idiopathic and symptomatic forms and may follow a chronic course for as long as forty years. Recently, a genetic factor has been identified in a group of patients with idiopathic narcolepsy (Yoss & Daly 1960). The idiopathic form usually appears during adolescence and is more common among males. The condition can be treated symptomatically with analeptic drugs.

**Insomnia.** Insomnia, or hypsomnia, is also associated with a variety of pathological states including brain tumor, metabolic disease, circulatory disorders, and aging. It is a common precursor of acute schizophrenic reactions and, of course, accompanies general emotional upset, depression, or anxiety. The hypsomnia may occur at the onset, during, or at the termination of sleep. Clinical observers have reported that difficulty in falling asleep is a common symptom in anxiety neuroses, whereas frequent and early awakening from sleep is a classic feature of depressive illnesses. Insomnia is usually a difficult treatment problem, although drugs, psychotherapy, and relaxation-training are often effective.



**Nocturnal behavioral symptoms.** Behavioral symptoms during sleep—including night terrors, sleep-walking, enuresis, grinding of teeth, delirium, and nightmares—can occur with specific disorders of the nervous system or as transient symptoms in normal individuals. All of these abnormalities may have some connection with aberrations of normal dreaming. Nearly all involve dreamlike activity and tend to occur during low-voltage EEG phases.

**Sleep therapy.** Sleep can be learned as a response in order to reduce anxiety. Since the excessive sleeper may use sleep to "narcotize" anxiety, it may be supposed that sleep would be useful as an adjunct to psychotherapy. Sleep therapy was first used in Switzerland, but it has achieved greatest popularity in the Soviet Union. As a consequence of Pavlov's theories, prolonged sleep is used to "protect" the brain during mental illness. It is said to be most effective in cases of catatonic excitement, depression, and acute anxiety, but no controlled experimental studies were located. [See MENTAL DISORDERS, TREATMENT OF, article on SOMATIC TREATMENT.]

**Conclusion.** Since the first comprehensive summary of sleep and wakefulness by Kleitman (1939) there has been a considerable increase in scientific interest in the problem. At that time there were about thirteen hundred references in the literature. There are now more than four thousand. Research activity is much greater, however, in neurophysiology than in psychology or other disciplines. Psychologists for the most part seem to have taken the transition from waking to sleep as a natural boundary for the study of behavior. Sleep, however, is not behaviorally empty. Even in deepest sleep subjects retain responsiveness to external stimuli. This finding that a considerable amount of behavior can be induced during sleep raises some interesting problems for both psychologists and neurophysiologists. If sleep is not, behaviorally, a silent state, what limits does sleep impose on the organization of behavior? If the pattern and meaning of auditory signals can be recognized by the sleeping subject, how are we to define sleep? How is the central nervous system organized during the sleeping state?

A few answers have been suggested to the fundamental questions raised by the identification of similar cyclic patterns of sleep in man and animals. The discovery of the physiological correlates of dreaming permitted the development of entirely new concepts relating animal to human subjective states. But as research increases, questions multiply. Do animals dream? Is the dreaming phase in man under the control of the limbic system? What organizing concept in neurophysiology will ac-

count for the simultaneous high activation levels and high response threshold seen in the emergent low-voltage stage of sleep?

The age-old questions of the causes of sleep, the biological functions of sleep, the neural and humoral mechanisms of sleep, the effects of distributed sleep, and the modifiability of the sleep-wake cycle are only partially answered. Clearly, research on the psychological, physiological, and social properties of sleep has important implications for all of the life sciences. There are practical reasons, too, for the brisk developments in this field. The requirements of space flight, of modern industry, and of military operations put the subjects of sleep and waking, and of biological rhythms in general, into challenging perspective.

HAROLD L. WILLIAMS

[See also DREAMS and FATIGUE. Other relevant material may be found in HYPNOSIS; MENTAL DISORDERS, article on BIOLOGICAL ASPECTS; NERVOUS SYSTEM, articles on STRUCTURE and FUNCTION OF THE BRAIN and ELECTROENCEPHALOGRAPHY.]

#### BIBLIOGRAPHY

- BARTLEY, SAMUEL H.; and CHUTE, E. 1947 *Fatigue and Impairment in Man*. New York: McGraw-Hill.
- BILLS, ARTHUR G. 1931 Blocking: A New Principle of Mental Fatigue. *American Journal of Psychology* 43: 230-245.
- CIBA FOUNDATION FOR THE PROMOTION OF INTERNATIONAL COOPERATION IN MEDICAL AND CHEMICAL RESEARCH 1961 *Symposium on the Nature of Sleep*. Edited by G. E. W. Wolstenholme and Maeve O'Connor. Boston: Little.
- COUNCIL FOR INTERNATIONAL ORGANIZATIONS OF MEDICAL SCIENCES 1954 *Brain Mechanisms and Consciousness*. Proceedings of a conference held at Ste. Marguerite, Quebec, Canada, August, 1953. Edited by Edgar D. Adrian et al. Oxford: Blackwell; Springfield, Ill.: Thomas.
- HENRY FORD HOSPITAL, DETROIT 1958 *Reticular Formation of the Brain: Symposium*. Edited by Herbert H. Jasper et al. Boston: Little.
- JOUVET, M. 1961 Telencephalic and Rhombencephalic Sleep in the Cat. Pages 188-208 in Ciba Foundation, *Symposium on the Nature of Sleep*. Edited by G. E. W. Wolstenholme and Maeve O'Connor. Boston: Little.
- KLEITMAN, NATHANIEL (1939) 1963 *Sleep and Wakefulness*. Rev. & enl. ed. Univ. of Chicago Press.
- LINDSLEY, DONALD B. 1960 Attention, Consciousness, Sleep and Wakefulness. Volume 3, pages 1553-1593 in *Handbook of Physiology*. Section 1: Neurophysiology. Edited by H. W. Magoun et al. Baltimore: Williams & Wilkins.
- LOBASHEV, M. E.; and SAVVATEEV, V. B. 1959 *Fiziologiya sutochnogo ritma zhivotnykh* (Physiology of Circadian Rhythms in Animals). Moscow and Leningrad: Izdatel'stvo Akademii Nauk SSSR.
- LUBY, ELLIOT D. et al 1960 Sleep Deprivation: Effects on Behavior, Thinking, Motor Performance and Biological Energy Transfer Systems. *Psychosomatic Medicine* 22:182-192.

- OSWALD, IAN 1962 *Sleeping and Waking: Physiology and Psychology*. Amsterdam and New York: Elsevier.
- PIÉRON, HENRI 1913 *Le problème physiologique du sommeil*. Paris: Masson.
- SIMON, CHARLES W.; and EMMONS, WILLIAM H. 1955 Learning During Sleep? *Psychological Bulletin* 52: 328-343.
- WILKINSON, ROBERT T. 1963 Interaction of Noise With Knowledge of Results and Sleep Deprivation. *Journal of Experimental Psychology* 66:332-337.
- WILLIAMS, HAROLD L.; LUBIN, A.; and GOODNOW, J. J. 1959 Impaired Performance With Acute Sleep Loss. *Psychological Monographs* 73, no. 14:1-26.
- WOLF, WILLIAM (editor) 1962 *Rhythmic Functions in the Living Systems*. New York Academy of Sciences, *Annals* 98:753-1326.
- YOSS, ROBERT E.; and DALY, DAVID D. 1960 Hereditary Aspects of Narcolepsy. *American Neurological Association, Transactions* [1960]:239-240.

### SLUTSKY, EUGEN

Eugen Slutsky, or Evgenii Evgenievich Slutskii (1880-1948), Russian economist, statistician, and mathematician, was born in Yaroslavl province, the son of a schoolteacher. He entered the University of Kiev as a student of mathematics in 1899 but was expelled three years later for revolutionary activities. From 1903 to 1905 he studied engineering at the Institute of Technology in Munich. After the revolution of 1905 he returned to Russia, obtaining a degree in law from the University of Kiev in 1911. For a while he taught jurisprudence in a technical college, but he became interested in political economy and in 1918 he received a degree in that subject and became a professor at the Kiev Institute of Commerce. In 1926 he left Kiev for Moscow to join the Kon'iunkturnyi Institut (i.e., the Institute for the Study of Business Cycles). From 1931 to 1934 he was on the staff of the Central Institute of Meteorology. Moscow State University conferred an honorary degree in mathematics on him in 1934. From that time until his death he held an appointment at the Mathematical Institute of the Academy of Sciences of the Soviet Union.

One of Slutsky's first published works was his famous article on the theory of consumer behavior (1915). Slutsky here developed some ideas of F. Y. Edgeworth's and Vilfredo Pareto's on the relationship between the utility function and prices, income, and consumption. His main achievement was to show that, with money income fixed, any change in the price of a commodity can be divided into two parts. The first part is the change in relative prices, with real (not money) income fixed. This is called the substitution effect; the consumer maintains approximately a given indifference level.

The second part is the balance of the price change (a proportional shift in all prices), which can be translated into an equivalent change in money income, with prices constant, causing a variation in real income. This is called the income effect; the consumer shifts from one indifference level to another. The two effects turn out to be independent and additive ("Slutsky's relation").

The distinction drawn between the substitution effect and the income effect is often left ambiguous. There is a choice as to whether to define the holding of real income fixed as (a) holding the consumer's utility fixed, or (b) holding a Lespeyre measure of the purchasing power of his money income fixed. Slutsky clearly uses (a) in his mathematical analysis and (b) in his verbal explanation. The difference is, however, of the second order of smalls and unimportant when considering small changes in price. (For a further discussion of this point, see Mosak 1942.)

Although in R. D. G. Allen's opinion (1950) the present theory of consumer behavior is as much a development of Slutsky's work as of Pareto's, Slutsky's paper attracted no attention until Allen discovered it in the mid-1930s (1936).

In his later years Slutsky did very little work in pure economics but made a considerable contribution to mathematical statistics and the theory of probability. His first paper on this subject, "On the Criterion of Goodness of Fit of the Regression Lines and on the Best Method of Fitting Them to the Data" (1913), was written eight years before the work of R. A. Fisher on the same subject. Slutsky was interested in the problem of formalizing the theory of probability. In the article "Über stochastische Asymptoten und Grenzwerte" (1925) he examined the accurate definition of the concept of asymptotic convergence in probability, and so made an important contribution to the theory of probability (see also 1928; 1938).

Slutsky was one of the originators of the theory of stochastic processes. In his work "The Summation of Random Causes as the Source of Cyclic Processes" (1927) he proved that the "periodical" oscillations in economic, meteorological, and other time series do not necessarily show the presence of any underlying periodic cause; such oscillations are typical of all serially correlated random sequences, including those which result from taking a moving sum in a sequence of mutually independent, purely random quantities. By repeatedly smoothing such sequences, one can obtain a new sequence, which will in any limited period of time be closely approximated (with probability one) by a sine curve. This effect of averaging random



series was studied by Slutsky independently of G. U. Yule.

Slutsky's study of stochastic processes centered on serial, or lag, correlation, that is, correlation between two members of a time series separated by a fixed time interval. In many important applications serial correlation depends only on interval length, not on position along the time series. The concept of serial correlation has broad applications in many branches of science and technology, including radio engineering. Slutsky's theory was generalized by A. Khinchin (1934).

Slutsky wrote several articles on the estimation of parameters of stochastic processes. His complete works on the theory of probability and mathematical statistics have been published (see *Izbrannye trudy*).

In the last years of his life Slutsky studied the problem of computing tables for functions of several variables. This work resulted in the posthumous publication of *Tablitsy dlia vychisleniia nepolnoi  $\Gamma$ -funktsii i funktsii veroiatnosti  $\chi^2$*  (Tables for the Computation of the Incomplete  $\Gamma$ -function and the  $\chi^2$  Probability Distribution; 1950).

A. A. KONÜS

[For the historical context of Slutsky's work, see the biographies of EDGEWORTH and PARETO; for discussion of the subsequent development of Slutsky's ideas, see TIME SERIES; UTILITY.]

#### WORKS BY SLUTSKY

- 1913 On the Criterion of Goodness of Fit of the Regression Lines and on the Best Method of Fitting Them to the Data. *Journal of the Royal Statistical Society* 77:78-84. → Published in Russian in 1960 in *Izbrannye trudy* (Selected Works).
- 1915 Sulla teoria del bilancio del consumatore. *Giornale degli economisti* 3d Series 51:1-26. → Published in Russian in 1963 by the Akademiia Nauk SSSR, in *Ekonomiko-matematicheskie metody*, Volume 1.
- 1925 Über stochastische Asymptoten und Grenzwerte. *Metron* 5, no. 3:3-89. → Published in Russian in 1960 in *Izbrannye trudy*.
- (1927) 1937 The Summation of Random Causes as the Source of Cyclic Processes. *Econometrica* 5:105-146. → First published in Russian. Reprinted in 1960 in *Izbrannye trudy*.
- 1928 Sur les fonctions éventuelles continuées, intégrables et dérivables dans le sens stochastique. *Académie des Sciences, Paris, Comptes rendus hebdomadaires des séances* 187:878-880. → Published in Russian in 1960 in *Izbrannye trudy*.
- 1938 Sur les fonctions aléatoires presque périodiques et sur la décomposition des fonctions aléatoires stationnaires en composantes. Pages 33-55 in *Colloque consacré à la théorie des probabilités*. Part 5: Les fonctions aléatoires, by S. Bernstein, E. Slutsky, and H. Steinhaus. *Actualités scientifiques et industrielles*, No. 738. Paris: Hermann. → Published in Russian in 1960 in *Izbrannye trudy*.

1950 *Tablitsy dlia vychisleniia nepolnoi  $\Gamma$ -funktsii i funktsii veroiatnosti  $\chi^2$* . Edited by A. N. Kolmogorov. Moscow: Akademiia Nauk SSSR. → Published posthumously.

*Izbrannye trudy: Teoriia veroiatnostei, matematicheskaia statistika*. Moscow: Akademiia Nauk SSSR, 1960. → Contains papers originally published between 1913 and 1938.

#### SUPPLEMENTARY BIBLIOGRAPHY

- ALLEN, R. D. G. 1936 Professor Slutsky's Theory of Consumers' Choice. *Review of Economic Studies* 3:120-129. → A summary of Slutsky's 1915 article on consumer behavior.
- ALLEN, R. D. G. 1950 The Work of Eugen Slutsky. *Econometrica* 18:209-216. → Includes a bibliography on pages 214-216.
- KHINCHIN, A. 1934 Korrelationstheorie der stationären stochastischen Prozesse. *Mathematische Annalen* 109: 604-615.
- MOSAK, JACOB 1942 On the Interpretation of the Fundamental Equation of Value Theory. Pages 69-74 in *Chicago, University of, Department of Economics, Studies in Mathematical Economics and Econometrics in Memory of Henry Schultz*. Edited by Oskar Lange, Francis McIntyre, and Theodore O. Yntema. Univ. of Chicago Press.

SMALL, ALBION W.

Albion Woodbury Small (1854-1926) did more than any other American sociologist to establish the recognition of sociology as an academic subject, and he shared with Lester F. Ward and Franklin H. Giddings the leading role in defining the scope and status of sociology among the social sciences.

Like many sociologists at the turn of the century, Small was initially trained in theology, but in his case that training had been greatly broadened and secularized before he began to teach sociology. After graduating from Colby College in 1876 and from the Baptist Newton Theological Seminary in 1879, he spent two years studying at Berlin and Leipzig. There he was deeply influenced by such historical economists as Gustav Schmoller and by such welfare economists as Adolf Wagner and Albert Schäffle. He read Karl Marx with sympathy but never became a convert. After returning to America, Small taught history and political science for several years at Colby before obtaining a doctorate in 1889 from the Johns Hopkins University. There he specialized in history and political science but was impressed by welfare economics as taught by Richard T. Ely; sociology was not offered at Hopkins at that time. The influence of the welfare economists and admiration for Ward led Small to find the ultimate justifica-

tion of sociology in its stimulation and guidance of sound social planning and in its development of a reliable body of secular social ethics.

**Establishing sociology as a profession.** In 1889 Small became president of Colby and remained there until 1892, when he was appointed to the first chair of sociology in the United States, at the newly founded University of Chicago. There he built the first, and long the best, department of sociology in the United States, appointing in the early years George E. Vincent, W. I. Thomas, Charles R. Henderson, and, somewhat later, Robert E. Park and Ernest W. Burgess. With Vincent, Small wrote the first sociology textbook, published in 1894. Continuing as chairman of the sociology department until his retirement in 1925, Small also served as dean of the graduate school after 1905. Throughout his academic career, he was respected not only as an administrator but also as a conscientious and informative teacher.

Small founded the *American Journal of Sociology* in 1895 and edited it for the next 30 years. It was long the only substantial sociological journal in the United States and for that time the best in the world. Through this medium Small exerted a dominant influence in extending academic and public interest in sociology as well as in raising the standards for sociological literature. He did much to familiarize young scholars with German sociology, an ever more important service as the number of Americans studying for their doctorates in Germany decreased. Small helped to establish the American Sociological Society in 1905 and served as its president from 1912 to 1913. Until his retirement he was one of the most energetic leaders of the society, assuming the task of editing and publishing the *Annual Proceedings* of its meetings. He was also prominent in international sociological activities and served as president of the Institut International de Sociologie in 1922.

**Contributions to sociology as a discipline.** Although Small may ultimately be remembered for his lifelong campaign to establish sociology as a valid field of academic endeavor rather than for his intellectual contributions, the development of his ideas nevertheless exemplifies a general trend in sociology during his lifetime. This trend, as Small assessed it, was a transition from primary attention on relatively static social structures to a dynamic and functional analysis of social processes. In his 1894 textbook he had shown much interest in the social organism analogies that were so popular in the early period of sociology, but through his later work he became by far the most important American figure in promoting the recog-

nition and elucidation of the concept of the social process.

The basic raw material of the social process is, according to Small, group activities. Group activities tend to be based on elemental human interests, and the inevitable conflict of these interests provides the dynamics of the social process. This conception of Small's was first shaped by his reading of Marx and by his studies under Wagner and Schäffle which emphasized the economic pressures that produce social reconstruction. Small had also read the earlier books of Ludwig Gumplowicz, which stressed the conflict of social groups, but was repelled by Gumplowicz's naturalistic evolutionism, which maintained that mankind could not provide for the betterment of social conditions. Small believed that conflicts could be accommodated and anarchy prevented if the conflicts were carried out under the authoritative supervision of the state, which adjudicates group antagonisms.

As early as 1893 Small had constructed a comprehensive schedule of the human interests that emerge in comparable forms of group manifestations. While he was still developing this idea, he came into contact with *Wesen und Zweck der Politik* (1893), by Gustav Ratzenhofer, the learned Austrian general who was also a social philosopher. Small's attempt to fuse his own views with those of Ratzenhofer was already evident by 1903, and when he published his *General Sociology* (1905), the fusion was virtually complete. Although *General Sociology* is widely considered to be the mature expression of Small's doctrine of the dynamic role of the conflict of interests, many of his students contend that he presented the subject in a more complete and intriguing manner in his famous lectures on "The Conflict of Classes," which, unfortunately, were never published.

Small's contribution to sociological methodology, which he considered vitally important, has come to be virtually ignored, sociology has since turned toward a more quantitative approach. He was very concerned with defining the scope and objectives of sociology, outlining the main subdivisions of the field and stating its fundamental ethical goals.

Small made significant contributions to social science outside the field of formal sociology, especially to political science and economics. His conception of the state as the mediator of conflicting group interests inspired Arthur F. Bentley's famous *The Process of Government*, perhaps the most important American contribution to political theory in the twentieth century. Several of Small's books contributed to the development of institutional economics (1907; 1909; 1913; 1924). His



vigorous critique of the capitalist system was influenced not only by Marxism and the early welfare economists but also by his (temporary) Chicago colleague Thorstein Veblen and even more by Werner Sombart. It is not surprising that Small's wide-ranging acquaintance with the various social sciences should have led him, toward the end of his career, to be increasingly interested in promoting and encouraging their synthesis.

HARRY ELMER BARNES

[For the historical context of Small's work, see ECONOMIC THOUGHT, article on THE INSTITUTIONAL SCHOOL; SOCIAL PROBLEMS; and the biographies of ELY; GUMFLOWICZ; MARX; RATZENHOFER; SCHMOLLER; SOMBART; VELEN; WAGNER; WARD, LESTER F. For discussion of the subsequent development of his ideas, see the biographies of BENTLEY; BURGESS; OGBURN.]

#### WORKS BY SMALL

- 1894 SMALL, ALBION W.; and VINCENT, GEORGE E. *An Introduction to the Study of Society*. New York: American Book Company.
- 1905 *General Sociology: An Exposition of the Main Development of Sociological Theory From Spencer to Ratzenhofer*. Univ. of Chicago Press.
- 1907 *Adam Smith and Modern Sociology: A Study in the Methodology of the Social Sciences*. Univ. of Chicago Press.
- 1909 *The Cameralists: The Pioneers of German Social Polity*. Univ. of Chicago Press.
- 1910 *The Meaning of Social Science*. Univ. of Chicago Press.
- 1913 *Between Eras From Capitalism to Democracy*. Kansas City, Mo.: Inter-collegiate Press.
- 1916 *Fifty Years of Sociology in the United States (1865-1915)*. *American Journal of Sociology* 21:721-864.
- 1924 *Origins of Sociology*. Univ. of Chicago Press.

#### SUPPLEMENTARY BIBLIOGRAPHY

- BARNES, HARRY E. 1926 The Place of Albion Woodbury Small in Modern Sociology. *American Journal of Sociology* 32:15-44.
- BARNES, HARRY E. 1948 Albion Woodbury Small: Promoter of American Sociology and Expositor of Social Interests. Pages 766-792 in Harry E. Barnes (editor), *An Introduction to the History of Sociology*. Univ. of Chicago Press.
- GOODSPEED, THOMAS W. 1926 Albion Woodbury Small. *American Journal of Sociology* 32:1-14.
- HAYES, EDWARD C. 1927 Albion Woodbury Small. Pages 149-187 in Howard W. Odum (editor), *American Masters of Social Science*. New York: Holt.
- HOUSE, FLOYD N. 1926 A List of the More Important Published Writings of Albion Woodbury Small. *American Journal of Sociology* 32:49-58.
- MACLEAN, ANNIE M. 1926 Albion Woodbury Small: An Appreciation. *American Journal of Sociology* 32:45-48.
- RATZENHOFER, GUSTAV 1893 *Wesen und Zweck der Politik*. 3 vols. Leipzig: Brockhaus. → Volume 1: *Die sociologische Grundlage*. Volume 2: *Die Staatspolitik nach aussen*. Volume 3: *Der Zweck der Politik im allgemeinen*.

## SMALL GROUPS

See GROUPS.

## SMELL

See TASTE AND SMELL.

## SMITH, ADAM

Adam Smith (1723-1790) was born in Kirkcaldy, Fifeshire, a fisheries and mining town near Edinburgh. He was the son, by a second marriage, of Adam Smith, comptroller of the customs at Kirkcaldy, who died early in 1723. His mother, Margaret Douglas, was the daughter of a substantial landowner in Fifeshire. Smith lived with his mother whenever he was in Scotland until her death in 1784; he was her only child, and he remained a bachelor until his death.

Smith received his elementary schooling in Kirkcaldy and entered the University of Glasgow in 1737, graduating with an M.A. in 1740. He then went to Oxford University as a Snell fellow at Balliol College, where he stayed until 1746. Beyond the fact that at Glasgow he was a student of Francis Hutcheson, almost nothing is reliably known about his intellectual experiences either at Glasgow or at Oxford. From 1746 to 1748 he lived with his mother in Kirkcaldy, presumably continuing his studies and awaiting an opening for a career in some remunerative post. Between 1748 and 1751, under the sponsorship of some of the leading intellectuals of Edinburgh, he gave several successful series of public lectures, on rhetoric and belles-lettres, on jurisprudence, and perhaps on other subjects. On the strength of the reputation gained by these lectures, he was elected in 1751 to the professorship of logic at the University of Glasgow. When the chair of moral philosophy became vacant later in the same year, he was elected to that superior post, which he occupied until 1763. In 1759 he published his first book, *The Theory of Moral Sentiments*.

Early in 1764 Smith went to France as tutor of the young duke of Buccleuch, stepson and ward of Charles Townshend. Smith remained in France from early in 1764 until late in 1766, most of the time in Toulouse but for some months in Paris, where he saw a good deal of the leading physiocrats and philosophes. He also visited Geneva, where he made the acquaintance of Voltaire. After his return to England in 1766, he was until early in 1767 an adviser to Charles Townshend, then chancellor of the exchequer and working on his fatal plan for taxing the American colonies;

Smith's contribution to this plan, if any, is unknown. Endowed with a generous pension for life from the duke of Buccleuch, Smith returned in 1767 to Kirkcaldy, where he remained until early in 1773, working on *The Wealth of Nations*. From 1773 until early in 1776 he was again in London, completing the book but also advising the government occasionally on economic matters. On March 9, 1776, *The Wealth of Nations* was finally published, and soon thereafter Smith returned once more to Kirkcaldy. Early in 1778, he was appointed a commissioner of customs for Scotland and also a commissioner of the salt duties. These were not sinecure posts, as has often been alleged, but required his presence in Edinburgh for the greater part of each week throughout the year. For the rest of his life, he held these posts and lived in Edinburgh, where he died on July 17, 1790.

### Overview of writings

*The Theory of Moral Sentiments* and *The Wealth of Nations* were the only full-length books that Smith wrote; however, he kept revising both of them for successive editions, and the additions in the sixth edition of *The Theory of Moral Sentiments*, which was published only a few weeks before his death, were substantial in extent and importance. He contributed three essays to periodicals in 1755 and 1761, and a collection of essays on literary and philosophical subjects was first published posthumously in 1795 ("Essays on Philosophical Subjects" 1963). In 1896 Edwin Cannan edited and published a recently found student's report of Smith's economic lectures as given in 1763 at the University of Glasgow (*Lectures on Justice* . . . 1964). In 1963 John M. Lothian edited and published a student's report on Smith's lectures on rhetoric and belles-lettres at Glasgow in 1762/1763 (*Lectures on Rhetoric* . . . 1963), which he had shortly before purchased at an auction sale in Scotland. W. S. Howell, in an as yet unpublished manuscript, describes these lectures as "an important and original revolutionary document in an important revolution in the history of rhetorical theory," the "revolution" consisting of the substitution for the old rhetoric stemming from Aristotle and Cicero of a new rhetoric based on the new learning of Bacon, Descartes, Locke, and others. At the same auction Lothian also purchased a student's report of Smith's economic lectures at Glasgow. This report, which is now in the possession of the University of Glasgow, is substantially fuller and also, it has been stated, superior in quality to that published by Cannan in 1896. Smith is said to have lectured also at the University of Glasgow on nat-

ural theology. Shortly before his death he supervised the burning of almost all of his manuscripts, some 16 folio volumes.

Smith clearly had a wide range of interests. The evidence available suggests that he reached his basic methodological and philosophical principles early in his career and that his destruction of manuscripts before his death was probably motivated much more by dissatisfaction with their form or with their incomplete state than by any fundamental change in his views.

Almost everything Smith wrote, in its methodological implications if not in its concrete subject matter, has some relevance for social thought, but it is expedient here to concentrate on the two books he published during his lifetime. Many writers, including the present author at an early stage of his study of Smith, have found these two works in some measure basically inconsistent. But in much of his writing Smith worked from what he called systems and what today would be called models. He was aware that "systems" are incomplete in the factors they take into account. Had he been able to complete his total system, he would probably have demonstrated that the apparent inconsistencies were often not real ones, but were merely the consequences of deliberate shifts from one partial model to another.

In a letter of November 1, 1785, to a French correspondent, Smith wrote that he had "... two other great works upon the anvil; the one is a sort of Philosophical History of all the different branches of Literature, of Philosophy, Poetry and Eloquence; the other is a sort of theory and History of Law and Government. . . . But the indolence of old age, tho' I struggle violently against it, I feel coming fast upon me, and whether I shall ever be able to finish either is extremely uncertain" ([1785] 1896, p. 166). The manuscripts of these two "great works" presumably were among those destroyed shortly before his death.

### "The Theory of Moral Sentiments"

The first thing to note about *The Theory of Moral Sentiments* is its title. It is a "theory" or "system," that is, it consciously and deliberately employs some measure of patterned abstraction and thus does not profess to account for all the relevant facts of the real world. Its primary concern is only with that part of human psychology which is involved in the interrelationships of men living in communities—the "moral sentiments," that is, the passions, propensities, affections, feelings, whether of approbation or of disapproval, aroused by these interrelationships. These senti-



ments are intermediate, in degree of reflection or "reason" involved, between the basic instincts that man shares with the animals and the calculation or ratiocination of sophisticated man as a reasoning being. When and how and in what degree these sentiments operate Smith discovered through observation of his neighbors, and presumably also, although he apparently never explicitly said so, through disciplined introspection exercised on the assumption that men are substantially alike in their subrational psychology.

To show how the sentiments operate to socialize the individual, to fashion him into a disciplined member of a harmonious social group, Smith introduced into his model the concept of "spectators," distinguishing two main species. There is, first, the spectator external to yourself, the "real" spectator, who, by manifesting in some manner his sentiment of approval or disapproval of your behavior, exercises an influence on you. There is, second, the internal spectator, yourself, operating on two distinct levels: first, your imagination of what the reaction of a hypothetical external spectator would be to your actual or contemplated behavior, second, your own moral judgment, the judgment by your own conscience, by "the man within your breast," by the "impartial spectator." This whole complex mechanism of psychological response by men to their neighbors' feelings of approval or disapproval, which Smith called sympathy, he regarded as the major factor in creating and maintaining a socialized community. It involves, according to Smith, not only the desire to win the praise or approval of others but also the desire to be praiseworthy; when the two desires are in conflict, conscience decrees that the latter shall prevail.

Commentators have objected that Smith here described a circular process, operating through sympathy like a set of "mirrors," and that he failed to explain adequately either what behavior is approved and what is disapproved or the origin or genesis of the social passions. Yet, at least by implication, Smith did offer such an explanation. He emphatically rejected human reason as the source of these sentiments. Also by implication he denied that there is a natural evolutionary process in which groups with a pattern of sentiments that is predominantly useful survive, whereas those groups with a pattern of antisocial sentiments perish. Smith maintained that man is endowed by God with his moral sentiments and that these sentiments bind men to each other because the deity so made them in its concern for the happiness of mankind. Smith ridiculed those who attributed to man's wisdom what is really the wisdom of God,

or of nature. Here Smith was, of course, invoking "final causes," or "the invisible hand."

It is hard for some people today to believe that Smith's optimistic deism was completely sincere, and they tend to attribute his exposition of it to prudential considerations or to concessions to a mode of speech called for by the standards of propriety of the time. But in the "enlightened" Scottish circles of Smith's time optimistic deism, sincerely held, was practically universal. Although orthodox Calvinists rejected its optimistic aspect as not religious enough and David Hume rejected it as calling for too much religious faith, aside from Hume, no one among Smith's teachers, colleagues, friends, or followers is identifiable as a critic of optimistic deism.

There may be genuine difficulty in reconciling Smith's deistic interpretation of the origin of the moral sentiments with other aspects of his social thought, including some of his specifically economic thought. Smith attributed to providence the original endowment of mankind with a set of moral sentiments conducive to the happiness of mankind. But unless he also assumed that providence intervenes constantly or intermittently to make appropriate adjustments in these sentiments as the physical or human environment changes through time (for which belief there is no evidence in anything he wrote), Smith would seem to have been postulating a static social psychology, at least on the subrational level, in what he himself admitted to be in many relevant respects a constantly changing and evolving world. Smith did recognize the impact of the variability of custom and fashion on the mode of operation of the moral sentiments, but, in spite of this variability, he specifically recognized only one major historical affront to the system of moral sentiments—the prevalence of infant exposure in the later period of ancient Greek civilization. He disposed of this exception as a temporary aberration, outweighed by the many outstanding virtues of Greek civilization at the time. Nowhere did he attempt to explain how antisocial passions and aberrations in conflict with the "Author of Nature's" design came into existence. But in keeping with the notion then held by some scientists that there is in nature a self-equilibrating mechanism by which aberrations are prevented from prevailing, a notion having some analogy to the modern scientific notion of "homeostasis," Smith held that there is an inherent tendency in the moral sentiments to overcome such aberrations.

Smith's treatment of "justice" in *The Theory of Moral Sentiments* is especially important for a proper interpretation of *The Wealth of Nations*.

Smith always used the word to mean substantially what Aristotle and the Schoolmen meant by "commutative justice." Justice is a negative virtue; it consists of refraining from injury to another person and from taking or withholding from another what belongs to him. It is thus distinct from benevolence, friendship, or charity. Smith considered justice, so understood, to be the necessary foundation of a viable society. It is a moral sentiment and thus finds voluntary or natural expression. The natural or spontaneous sentiment of justice is not, however, strong enough in ordinary men to meet the needs of society. Consequently, men have been endowed with the propensity to formulate rules of justice on the basis of their experience and reason, and they accept these rules for themselves and press them upon others. But even this is inadequate for the needs of society, and therefore government is established, its chief function being the coercive enforcement of justice on the individual members of the community through law and the magistrates.

The moral sentiments operate at different levels of intensity according to the nature and the strength of the external stimuli impinging upon men. Smith's discussion here closely parallels Hume's discussion in *A Treatise of Human Nature* (1739-1740, book 2, part 2, especially sec. 4) and elsewhere: in describing the way in which the strength of the "passions" between individuals varies with the closeness of their relationship with respect to duration, space, kinship, nationality, occupation, rank, and so forth, Hume repeatedly used the term "distance" metaphorically to signify any factors separating individuals from each other, a usage that goes back at least to Aristotle and Thomas Aquinas. Although Smith stressed distance in its primary spatial sense as an important factor in weakening the intensity of the moral sentiments—"To what purpose should we trouble ourselves about the world in the moon? All men, even those at the greatest distance, are no doubt entitled to our good wishes, and our good wishes we naturally give them. But if, notwithstanding, they should be unfortunate, to give ourselves any anxiety upon that account seems to be no part of our duty" ([1759] 1966, p. 197)—he also used the same idea, if not the actual term "distance," for the absence not only of spatial proximity but also of membership in the same family, village, town, province, country, circle of friends, guild or company, church, social class, or some other psychologically unifying bond. In similar manner he took it for granted that the participants in a large number of the transactions which occur in the market are (in the metaphorical sense) at an extreme dis-

tance from each other; they are, in relation to each other, anonymous, or strangers, so that there is limited occasion for any moral sentiments other than justice to come into operation.

In *The Theory of Moral Sentiments*, Smith minimized the contribution that even the highly successful pursuit of wealth or of higher social status makes to the happiness of an individual. Both in this work and in *The Wealth of Nations* Smith treated increase in aggregate wealth as a highly worthy objective for a country, but apparently in only one passage in either work is increase in *per capita* wealth or income expressly mentioned as a reason for the advantage of an increase in aggregate wealth. Smith attached little importance to an increase above a quite modest level of *per capita* income, but he attributed great value to the increase of population that an increase in aggregate wealth fosters and supports. With his optimistic view of the amount of happiness ordinarily enjoyed even by the poor, Smith believed that growth of population is ordinarily conducive to growth in the aggregate amount of human happiness. He also found value in increase of aggregate wealth because it makes possible an increase in handsome buildings and great avenues in the towns, the "magnificence" so extolled by the writers of classical antiquity and of the Renaissance, but he treated these as public rather than individual riches. Smith also included as an advantage of growth of aggregate wealth the progress of aesthetic and intellectual culture and of "civilization" in general, which he associated with communal enrichment.

### "The Wealth of Nations"

*The Wealth of Nations* is of great importance for three main reasons. First, it presents an impressive collection of economic data, gathered together by Smith from wide reading in publications from the time of the ancient Greeks and Romans to his own time and from acute observation of Britain and France in his own time. Smith used this material to illustrate and support his analysis of contemporary economic process, to provide a factual basis for his frequent ventures into the philosophy or theory of economic development, and at times to digress into narrative history presented without any clear relationship to his theoretical endeavors. *The Wealth of Nations* was heavily drawn on as a reliable source book for factual data by several generations of writers on economic matters, especially, of course, in the English-speaking countries. It is still useful for this purpose.

Second, it was the most comprehensive and ambitious attempt up to Smith's time to present in



comprehensive, and at the same time coordinated, fashion the nature of economic process in a predominantly "individualistic," or "competitive," or "market," or "capitalistic" society, to use modern adjectives. For a long time it largely determined the selection of issues and the initial analytical approach of economists in many countries, even when in their treatment of *The Wealth of Nations* they were much more critics than disciples.

Third, it was an evaluating and crusading book, which sharply criticized existing society and government and argued strongly for changes in national policy, especially in relation to the extent and nature of government intervention in economic matters—domestic, colonial, and international. Not immediately, but within a generation, it became a powerful influence on writers on economic policy. Later still, both directly and indirectly through those influenced by it, it became a significant factor in determining the course of national policy not only in Britain but in other countries as well. This is much more than any other economic work has ever achieved; and Smith probably has had much more influence than any other economist.

**Economic development.** Smith was deeply interested in the history, the causes, and the natural and artificial limitations of what we now commonly call "economic development" and what he referred to as "progress," "improvement," "progress of improvement," and "progress of opulence." His treatment of economic development is scattered throughout almost the whole work. He put most stress on the following factors as favorable to economic development: abundance of natural resources; technological progress as promoted by extension of division of labor; freedom of private enterprise from its own propensity to monopolistic organization; freedom from such hurtful artificial institutions as primogeniture; and freedom from official policies and practices that act as brakes on individual initiative or misdirect it. The basic source of economic progress, however, he found in the striving of individuals to improve their economic status or their rank in society—"... the desire of bettering our condition, a desire which ... comes with us from the womb, and never leaves us till we go into the grave" ([1776] 1950, p. 323). He did not believe that this desire does or should operate without restraint. It is disciplined by the sentiment of justice and by governmental enforcement of justice. It has to compete, beneficially or otherwise, with "the passion for present enjoyment," which acts as a restraint on accumulation; with indolence; with the occasional "liberality" of

employers to their workmen and of landlords to their tenants; and with "the pride of man [which] makes him love to domineer," so that a plantation owner may prefer the service of slaves to that of freemen even where the latter would be more profitable. All of these are factors that are present in a different degree in different orders of society and in different circumstances. Smith, moreover, saw the desire for individual enrichment and the desire to preserve or improve one's social status as occasionally coming into conflict, as, for instance, when one's rank calls for profuse expenditure but the preservation or augmentation of one's estate calls for frugality.

Smith placed great emphasis on the division of labor as a requisite of economic development; he also stressed the interrelations of the division of labor with technology and with commerce. He found a subrational or nonutilitarian origin for the resort to specialization (although not for its intensification) in a "propensity to truck and barter" innate in mankind. Here, following certain predecessors, he identified three contributions that the division of labor makes to productivity: by permitting indefinite repetition of simple tasks, it promotes dexterity; it eliminates the loss of working time involved in changing from one task to another; it facilitates invention of machinery, both by the artisans on the job and by outside observers. Smith stated that extension of the division of labor makes more capital necessary and therefore makes frugality and accumulation economic virtues. He always minimized the differences of innate ability or aptitude between different persons and thus gave little or no weight to the advantage claimed for specialization by a continuous chain of writers from ancient Greece on—that it makes possible the assignment of workers to those tasks for which they have the greatest aptitude. Smith pointed out that the division of labor is limited by the extent of the market and that growth of population both constitutes an extension of the market and is made possible by the increase of aggregate production that results from an extension of the division of labor. Growth of population, growth of aggregate wealth and income, and extension of the division of labor are thus explained as mutually dependent and mutually supporting factors.

Smith recognized, under the influence of Rousseau, that the division of labor has a drawback from a humanitarian point of view: the worker as a person tends to be degraded by the monotony of his work and the enlistment of only a narrow range of his mental faculties. Smith thought, however, that this predicament could be remedied by educa-

tion. This is one of the reasons that Smith accepted as a desirable function of government the financing, at public expense, of elementary education for the children of the poor. Smith's discussion was cited by Karl Marx in his presentation of the thesis that a degradation or "alienation" of labor is a consequence of division of labor, but Marx was unaware of the prior treatment of the subject by Rousseau to which Smith was indebted.

Smith's belief that the tendency to aggregate improvement is "natural," i.e., essentially the product of man's basic psychology, may have been a factor contributing to his skepticism about the possibility that government may make major positive contributions to economic development. While he charged government in general with operating as a brake on economic progress, Smith nevertheless remained an optimist. In man's zeal to better his condition, the "wisdom of nature" had provided a counterforce to mistaken government policies and practices that was sufficiently powerful to make possible, in most cases, a thriving and prospering economy.

**Smith's eclecticism.** It is a common error to interpret *The Wealth of Nations* as an unqualified eulogy of private enterprise and the businessman. It was only private enterprise operating in a fully competitive manner that Smith praised. He depicted businessmen in general as having a constant propensity to organize themselves into groups capable of exercising "monopoly" power, groups to which he indiscriminately attributed the capacity and, by implication, the will to exact the highest price at which any sales can be made. He also charged businessmen with major responsibility for persuading or pressing government to establish special privileges and legal monopolies for favored groups. Where monopoly is unavoidable, he preferred government to private operation. He had only deep and violently phrased scorn for the morals of businessmen organized in groups either to operate as monopolists or to obtain special privileges from government. *The Wealth of Nations* does lavish praise on the businessman, but only when he is on his good behavior.

Smith's main merits as an "analytical" or "scientific" theorist, to use modern eulogistic terms for "pure" economic theory, lie in his eclectic spirit. While deliberately resorting to abstraction, he very much doubted that abstraction could provide either understanding of the real world or, by itself, safe guidance for the legislator or statesman. On specific points of economic analysis some predecessors did better than Smith, and he failed to absorb fully some of the genuinely valuable analytical

contributions of Hume, the physiocrats, and Turgot. If "analytical" as a eulogistic term is to be interpreted strictly in terms of degree of rigor, internal consistency, and close analogy to abstract mathematical operations, Schumpeter's verdict that "the *Wealth of Nations* does not contain a single analytic idea, principle, or method that was entirely new in 1776" (Schumpeter 1954, p. 184) is difficult to challenge, and not merely because valuable ideas that are "entirely new" are hard to spot in any area of intellectual endeavor.

In both his major works Smith repeatedly amended his system, bringing into his discussion some hitherto neglected variable, some fresh observation of fact, some new objective. He has been rightly charged by critics with resorting profusely to such qualifications as "perhaps," "generally," and "in most cases," with the consequence that his models are not tight or rigorous. It is arguable, however, by those who, if forced to choose, prefer realism, or at least the pursuit of it, to rigor and elegance of analysis, that both of his major works are on the whole made better by the qualifications he sprinkled in their pages and that he would have made them still better, although still untidier, if he had used even more qualifying adjectives or phrases. He would at least have made it harder for later critics to use short quotations, out of context or stripped of their qualifications, to show his inability to avoid flagrant self-contradiction.

The question of the relation of relative labor input to exchange value is one instance where Smith appears repeatedly to have shifted from one belief to another; however, it may be that actually he was only shifting from one abstraction to another, while decorating his exposition (in a manner common then and not unknown now) with traditional maxims exalting the role of labor—maxims whose familiarity alone made them seem to carry logical or empirical weight. Smith can be quoted in support of all of the following propositions: that labor is the sole "source" of market value; that labor is the sole regulator of exchange value; that labor has, among the elements entering into production, a peculiar and perhaps even an exclusive value-creating power; that the relative values of different commodities are, or should be, proportional to their labor-time costs or to their wage costs; that all incomes are extracted from the product of labor. For some economists any one of these propositions suffices to label its exponent a "labor theory of value" theorist. It seems safer, nevertheless, not to attribute to Smith much more than the belief that in commercial or capitalist economies relative labor-time costs per unit of product have a large part in



the determination of the exchange values of different commodities, and relative wage-costs even more.

**Economic policy.** The separation of normative from non-normative, or policy, economics was a late development in the history of economics, and even today it is hard to fully execute this separation because many of the standard terms used in economic analysis carry with them an almost automatic normative or evaluative implication: for example, "productive," "utility," "value," "equilibrium." Prior to Smith's time it was rare for any writer to attempt to distinguish between, on the one hand, the study of economics in the purely "scientific" sense of the pursuit of understanding for its own sake, and, on the other hand, the use of economic analysis as an instrument for the formulation or evaluation of national economic policy. (Richard Cantillon was one of the few who did make this distinction.)

When Smith wrote *The Wealth of Nations* the term "political economy" was already in wide use. It was used with some ambiguity, but predominantly with emphasis on "political," indicating reference to national policy. The term "economics" was rarely used by itself except in its original Greek meaning of household management. When Smith chose as the title of his book *An Inquiry Into the Nature and Causes of the Wealth of Nations* instead of something like "Principles of Political Economy," it may be surmised that he did so because he thought of his book as including both an objective study of processes and causes, such as would be the subject matter of a treatise on physics or physiology, and a discussion of "political economy" proper, or an evaluative or hortatory treatment of governmental economic policy. Smith used the term "political economy" a dozen or so times, and every time, except perhaps once, he meant the economic policy of a nation. Since Smith generally took a dim view of the benefits to be derived from national economic policy, political economy must for him have been nearly synonymous with "economic poison."

Smith, of course, was not an exponent of philosophical anarchism, which apparently had nowhere been systematically expounded before William Godwin's *Enquiry Concerning Political Justice* (1793). If Smith had adopted the term "laissez-faire" as an appropriate label for his own policy views, he undoubtedly would not have interpreted it literally as a condemnation of all government interference with the activities of private individuals. He was as emphatic as he could be on the vital need for government enforcement of justice,

and there is evidence in *The Wealth of Nations* that he would have included in this function not only the formulation of rules of justice and the provision of machinery for the punishment of their infraction, but also the prevention of certain infractions by such enactments as standardization of weights and measures, requirements that commodities offered for sale be so stamped as to indicate their quality, and the establishment of building standards that would hinder individuals from subjecting others to the risk of fire or to other hazards to their property or their personal safety. Smith assigned to government the care of the defense of the community against foreign aggression or internal disorder and the levy of taxes to finance these activities. He also conceded to government the provision of those services needed by the community which could not practically be entrusted to private enterprise, because of the scale on which they had to be carried out or for other special reasons. On at least two issues—a "standing" or professional army versus a militia and the autonomy of the East India Company—Smith expressed a strong preference for governmental control in addition to or instead of private management. In general, where monopoly was unavoidable, he much preferred that it be under public rather than private control.

Nevertheless, it is as an exponent of free enterprise; free trade; noninterference of government in the individual's choice of occupation, residence, or investment; freedom for the individual to make his economic decisions of all kinds in response to the price movements of free and fully competitive markets—in short, of "economic liberalism" or "laissez-faire," as these terms were used in the nineteenth century—that Smith made his chief mark on the history of economics and on the economic and social history of the Western world.

These economic freedoms were to Smith "natural rights," essential constituents of the dignity of man. He also valued them from a utilitarian point of view, as giving maximum scope for incentives to industry and to efficiency. In the international sphere he saw in them the most solid factors working to bring peace between nations. Modern economists find Smith's arguments oversimplified and perhaps also too emotional and one-sided. But many of them still acknowledge a strong influence of his writings on their system of values and gladly continue to do homage to his name.

JACOB VINER

[For the historical context of Smith's work, see ECONOMIC THOUGHT, article on MERCANTILIST

THOUGHT; and the biographies of CANTILLON; HUME; MANDEVILLE; for discussion of the subsequent development of his ideas, see LAISSEZ-FAIRE; and the biographies of BASTIAT; LAUDERDALE; RICARDO; SAY.]

## WORKS BY SMITH

- (1759) 1966 *The Theory of Moral Sentiments*. New York: Kelley.
- (1776) 1950 *An Inquiry Into the Nature and Causes of the Wealth of Nations*. Edited by Edwin Cannan. London: Methuen. → A two-volume paperback edition was published in 1963 by Irwin.
- (1785) 1896 [Letter to the Duc de la Rochefoucauld.] *Economic Journal* 6:165-166.
- Essays on Philosophical Subjects. Volume 5, pages 49-399 in Adam Smith, *The Works of Adam Smith*. Aalen (Netherlands): Zeller, 1963.
- Lectures on Justice, Police, Revenue and Arms, Delivered in the University of Glasgow . . . Reported by a Student in 1763*. Edited by Edwin Cannan. New York: Kelley, 1964. → The Cannan edition was first published in 1896.
- Lectures on Rhetoric and Belles Lettres, Delivered in the University of Glasgow . . . Reported by a Student in 1762-63*. Edited by John M. Lothian. London and New York: Nelson, 1963.
- The Works of Adam Smith*. 5 vols. Aalen (Netherlands): Zeller, 1963.

## SUPPLEMENTARY BIBLIOGRAPHY

- Adam Smith, 1776-1926: *Lectures*. 1928 Univ. of Chicago Press. → Contains lectures by John Maurice Clark and others to commemorate the sesquicentennial of the publication of *The Wealth of Nations*.
- BITTERMANN, HENRY J. 1940 Adam Smith's Empiricism and the Law of Nature. *Journal of Political Economy* 48:487-520, 703-734.
- BONAR, JAMES (1894) 1932 *A Catalogue of the Library of Adam Smith*. 2d ed. London: Macmillan.
- GODWIN, WILLIAM (1793) 1946 *Enquiry Concerning Political Justice and Its Influence on Morals and Happiness*. 3d ed., rev. 3 vols. Univ. of Toronto Press.
- HARVARD UNIVERSITY, GRADUATE SCHOOL OF BUSINESS ADMINISTRATION, BAKER LIBRARY, KRESS LIBRARY OF BUSINESS AND ECONOMICS 1939 *The Vanderblue Memorial Collection of Smithiana*. Boston, Mass.: The Library.
- HUME, DAVID (1739-1740) 1958 *A Treatise of Human Nature*. Edited by L. A. Selby-Bigge. Oxford: Clarendon. → Reprinted from the original edition and edited with an analytical index.
- MACFIE, A. L. 1959 Adam Smith's *Moral Sentiments* as Foundation for His *Wealth of Nations*. *Oxford Economic Papers* New Series 11:209-228.
- MACFIE, A. L. 1961 Adam Smith's *Theory of Moral Sentiments*. *Scottish Journal of Political Economy* 8:12-27.
- RAE, JOHN (1895) 1965 *Life of Adam Smith*. With an introduction and guide by Jacob Viner. New York: Kelley.
- ROSENBERG, NATHAN 1965 Adam Smith on the Division of Labour: Two Views or One? *Economica* New Series 32:127-139.
- SCHUMPETER, JOSEPH A. (1954) 1960 *History of Economic Analysis*. Edited by E. B. Schumpeter. New York: Oxford Univ. Press.

- SCOTT, WILLIAM R. 1937 *Adam Smith as Student and Professor*. Glasgow: Jackson.
- SPENGLER, JOSEPH J. 1959 Adam Smith's Theory of Economic Growth. *Southern Economic Journal* 25:397-415; 26:1-12.

## SMITH, WILLIAM ROBERTSON

William Robertson Smith (1846-1894) was born in Scotland, the son of a distinguished scholar and minister in the Free Church of Scotland. The father instilled in his son a love for learning and for free critical inquiry, especially in the field of Biblical studies.

While Robertson Smith was a student at the University of Aberdeen his main interest shifted to mathematics and physics, subjects that he never wholly abandoned. His work in these subjects, published while he was still in his early twenties, was well thought of, and many years later he was considered as a candidate for a chair of mathematics at Glasgow. However, his first published paper, entitled "Prophecy and Personality" (1868), derived from his Biblical studies, and it was this field that subsequently took overwhelming precedence. After his early training at Aberdeen he went to the New College seminary at Edinburgh. In 1870, when he was only 24 years old, he was appointed to the chair of Hebrew and Old Testament exegesis at the Free Church College at Aberdeen, and five months after this appointment he was ordained.

For six of the 11 years of his stay at Aberdeen, wrangles with the general assembly over the issue of Biblical criticism consumed his time and sapped his energies. Some of his lectures had aroused suspicion, but the matter came to a head with the appearance of his article "Bible" (1875) in the ninth edition of the *Encyclopaedia Britannica* and of a review of this article, ascribed to A. H. Charteris, professor of Biblical criticism at the University of Edinburgh, which accused Robertson Smith of importing discredited views from the Continent and of tarnishing a theological professoriate. The celebrated "Robertson Smith affair" had begun; thereafter he was hounded by the Free church, and rather than accept the odious limitations on intellectual activity that were the condition of his continuing occupancy of the chair at Aberdeen, he stood his ground and in 1887 was removed from his post. The two succeeding years were devoted almost entirely to writing articles for the *Encyclopaedia Britannica* and to the enormous amount of editorial work he undertook as its chief editor—it is said that he read the entire edition.



During the struggle with the governing body of the Free church, Robertson Smith had been offered two appointments by Harvard University: first, the chair of Hebrew and Oriental languages and, later, the chair of ecclesiastical history. He did not accept either, but in 1889 he was offered the chair of Arabic at Cambridge and accepted with great pleasure. He found time amid his academic activities to make several visits to Egypt, and on one of these he accompanied the celebrated explorer Richard Burton on a minor expedition. He spent two months in Arabia; he traveled in Palestine, Syria, and the Libyan desert; and during another period of leave he went to Tunisia and Algeria. He made his way easily among the Arabs and recorded some of his observations, but he lacked the time and the training for any systematic studies of social life.

Several influences contributed to Robertson Smith's intellectual development. His home environment encouraged the study of Hebrew, Arabic, and the classics, and he shared the interest in history and in the nature of human progress that was characteristic of the general Scottish intellectual environment. The latter preoccupation had produced in the eighteenth century a group of philosophers who were deeply curious about the evolution of social institutions (Bryson 1945). More directly, Robertson Smith was influenced by his friend J. F. McLennan, a Scottish barrister and the author of *Primitive Marriage and Studies in Ancient History*. McLennan, in turn, was familiar with the work of Fustel de Coulanges who, in *The Ancient City*, had functionally related the development of religion (from its classical roots to the appearance of Christianity) to the kind of social units into which people were arranged, although he did give religion a sort of causal priority in producing social change. [See the *biographies of McLennan and Fustel de Coulanges*.] An examination of Robertson Smith's general theories of religion clearly reveals how profoundly these ideas influenced him.

The two works of Robertson Smith's that were important for the development of sociology are *Kinship and Marriage in Early Arabia* (1885) and *Lectures on the Religion of the Semites* (1889). Of the two, the analysis of religion has had the greater influence, but its full understanding requires a knowledge of some of the ideas about tribal society in the analytically weaker book on kinship. In fact, the two works go hand in hand. (The very first formulation of Robertson Smith's general views on religion appeared in the well-known article "Sacrifice" in the ninth edition of the *Encyclopaedia Britannica* in 1886, a few months after the study of kinship.)

### Analysis of religion

As a churchman and theologian, Robertson Smith had an immediate concern with religion. His writings were, in part at least, the solution to a personal problem, giving him a more satisfactory understanding of the essence of religion and a reason for his continued convictions. It is only in the light of his commitment to religion that his long and acrimonious struggle with the Scottish Free church makes sense, for his aim was to help revitalize religion, not to destroy it.

His penchant for looking at things sociologically led him first of all to the problem of a satisfactory analysis of beliefs. His point of analytical departure was to argue that the beliefs people hold and the general dogmas to which they subscribe are not acquired by individual effort—that is, they are not a matter of individual choice resulting from reasoned argument but are part of the general social apparatus with which man is endowed at birth. Religion is a part of an individual's collective representations, to use the phrase Durkheim later introduced into sociological vocabulary. In other words, Robertson Smith saw religion as "a part of the organised social life into which a man was born, and to which he conformed throughout his life in the same unconscious way in which men fall into any habitual practice of the society in which they live" ([1889] 1927, p. 21). Thus gods and their worship are taken for granted, just as other usages are accepted, and "if they [men] reasoned or speculated about them, they did so on the presupposition that the traditional usages were fixed things behind which their reasonings must not go, and which no reasoning could be allowed to overturn" (p. 21).

Early religion, according to Robertson Smith, had very little to do with beliefs; it was concerned with practices, and conformity was a matter of course. Just as in studies of political structure, an inquiry does not begin "by asking what is recorded of the first legislators, or what theories men advanced as to the reason of their institutions" (p. 21), but by seeking to understand political institutions and their effect on social relations, so too the study of religion must be directed at "the workings of religious institutions" and the way in which the relations of the worshipers are shaped. The analogy is not arbitrarily chosen, for, Robertson Smith contended, "the parallelism in ancient society between religious and political institutions is complete" (p. 20).

Robertson Smith believed that political institutions are older than political theories, and in the same way, religious institutions are older than re-

ligious theories. Rites are prior to beliefs, rationalization follows practice. There is evidence for priority of ritual in early forms of religion, which were constituted wholly of ritual and practical usage. Practices that persist do tend to acquire accretions of explanatory beliefs, but the meanings thus attached are inconsequential, vague, or contradictory, while the practices are fixed. More often than not, people are unaware of the doctrinal differences that divide sects in the contemporary world, and when doctrinal explanations of ritual acts are given at all, these may differ within the same religion and culture without any question of orthodoxy being raised; conspicuously less leeway is permitted in the carrying out of proper procedures. Robertson Smith felt that sociological inquiry would do well to begin not with a creed or with what individuals privately think but with overt ritual acts. Some sixty years after Robertson Smith had made this point with crystal clarity, Radcliffe-Brown was still laboring it thus: "My suggestion is that in attempting to understand a religion it is on its rites rather than beliefs that we should first concentrate our attention" (1952, p. 155).

Robertson Smith's study of ancient religion led him to conclude that it is external religious acts that give religion its principal social meaning. In ancient society, the rites of religion ensured conformity: privately, a man might believe what he liked; only by public acts of piety was it possible to judge whether he subscribed to religion. It was through his acts that the participant revealed his social identity, and for this identification his beliefs were virtually irrelevant. Further, acts and ritual were necessarily rigid because they showed public conformity and were testimony of a person's social outlook. Nonconformity, therefore, meant that someone was at odds with his social group. Ritual was the discipline that ensured the cohesion and continuity of social groups. Performance of rituals pledged continued membership in a social group; default was a public withdrawal. Excommunication was in fact the severance of ties to a social group and was therefore the same as outlawry. Religious allegiance and political connection were one and the same thing; a change of one was a change of both. Alter a man's political status and he is compelled to change his religion, "for a man's religion is part of his political connection" ([1889] 1927, p. 36).

Viewed in this light, ancient religion had both regulative and stimulative functions. It was regulative in the sense that the welfare of the individual could reach its optimum development only to the extent of his compatibility with his fellowmen. Individual goodness, it follows from this, was the

behavior that made for the general harmony of the community. By the same token, sin was an act that seriously disrupted this internal harmony and produced an impasse in relations, leading to the kind of stultification that could be remedied only by excommunication from the religious congregation and by outlawry from the social group. Religion, however, was not merely regulative in the sense that it compelled conformity; it was also stimulative, in that ritual was a repetitive statement of divine and human unity. Each and every religious act revived and consolidated the community of worshipers, giving it a consistency of purpose and a continuity of being. Religion was a social activity from which the good of the community was derived. "Religion did not exist for the saving of souls but for the preservation and welfare of society . . ." (*ibid.*, p. 29). Religion was based on these two elements—the regulative and the stimulative—rather than on personal problems of redemption.

In his writings on sacrifice, Robertson Smith further analyzed the nature of the congregation of worshipers, which he regarded as "the central problem of ancient religion . . ." (p. 27). He rejected the several theories of sacrifice—gift, tribute, covenant, expiation, propitiation; indeed, he could hardly do otherwise, given his acceptance of the premise that religion is a social activity, because this premise implies that the *intentions* of the sacrificer are relatively unimportant. Instead, Robertson Smith focused his attention on those who participate in the sacrificial act by partaking in the commensal meal. He rejected the view, recently endorsed by Evans-Pritchard, that the festal meal which succeeds a sacrifice "is not a sacramental meal but an ordinary commensal act of family or kin which, moreover, falls outside the sacrificial rites" (Evans-Pritchard 1956, p. 274). The greater part of one chapter in Robertson Smith's *Lectures on the Religion of the Semites* (especially pp. 272–282 in the 1927 edition) was devoted to the argument that the meal succeeding the slaughter was *not* a mere family affair but was eaten by a congregation of worshipers. Today also, as the present author's studies show, it is not permitted among the Bedouin Arabs for one family alone to consume the flesh of a slaughtered domesticated animal.

Robertson Smith pointed out that the meal following a sacrifice further differed from an ordinary meal in two ways. First, the flesh of the sacrificial victim was *meat with a special status*, different from the meat of game. (Again, this is a distinction drawn also by contemporary Bedouins. When game—gazelle, for example—is captured, the captor may distribute it among the households in a camp



as a matter of generosity, or he and his family may consume it entirely, running no more risk than the charge of meanness.) A commensal meal, therefore, is not simply an arrangement by which people who happen to be at hand join in eating but one at which attendance is obligatory because the meat is that of a sacrificial animal. Confusion sometimes arises around this point because all animals are customarily slaughtered in a prescribed manner. But ritual slaughter does nothing more than make it permissible to eat the flesh; it does not command a congregation and must not be confused with an act of sacrifice. Whether or not the flesh provides a commensal meal depends on the status of the animal and its relation to the person who slaughters it or who has it slaughtered.

Second, the special character of the commensal meal following a sacrifice was related to the link between the owner of the slaughtered animal and the victim: the victim was part of the sacrificer in a very intimate sense. This is also true among modern Bedouins: a camel or a sheep is part of a man's *risiq* (wealth) in the same way as are his children, and they are frequently and affectionately referred to by the same word. The host who slaughters one of his own animals is offering something of himself—a part of his "spiritual essence," which creates "a sort of spiritual bond," as Mauss put it—and all those who accept are thereby accepting a special relationship with the host, based on commensal nourishment. They become men who have joined together to share their economic resources for the common preservation of their lives.

Robertson Smith maintained that eating together constituted a bond among men equivalent to kinship. Participants in the meal, regardless of their consanguineous connections, were men bound by mutual obligation (another way of saying "kinship") sealed and covenanted in God. For kinship, as Robertson Smith saw it, was a matter not of consanguinity but of defined social relationships. Shrewdly, he pointed out that the tie that united the foster child to his foster mother and her kin was a tie of milk, not blood ([1889] 1927, p. 274): closeness of kinship was established not by birth but by nourishment. The sacrificial meal was one that a man shared with his commensals, "so that commensality can be thought of (1) as confirming or even (2) as constituting kinship in a very real sense" (p. 274).

Thus, in the early history of religion, the important relationship was not that of the individual man to a supernatural power, but that of "all the members of the community to a power that has the

good of the community at heart . . ." (p. 55). Communities were small, and the rivalry between them was expressed through the gods of the communities, just as the rivalry between neighboring villages in Spain is expressed through their local saints (p. 31). "Solidarity of the gods and their worshippers as part of one organic society" (p. 32) kept groups of people apart. As society advanced and institutions were elaborated, religious organization developed in matching likeness. With the fusion of groups into larger unities, political growth was marked by the appearance of a hierarchy of new gods: "A society and kinship of many gods began to be formed, on the model of the alliance or fusion of their respective worshippers" (p. 39). Church organization and political divisions were but two aspects of the same thing. This part of Robertson Smith's argument clearly shows the influence of Fustel de Coulanges, whose theme in *The Ancient City* was the parallel development of religious and political organizations.

Durkheim, whose writings on religion have had much more general influence than those of Robertson Smith, later argued that social groups possess stability only if a sentiment of unity exists, which must be given collective expression in emblematic form; and through this emblem, society worships itself. This view is an elaboration of what Robertson Smith wrote, for like Durkheim, he saw in the unity of the social group the core of religion. Since the whole edifice of religion is built up out of the unity that exists in groups, from the smallest to the largest, the idea of God is rooted in the local community and its growth is the elaboration of political organization.

This general idea was accepted among social scientists until Evans-Pritchard—first in an article on Nuer sacrifices (1954) and then in his book *Nuer Religion* (1956)—put forward a fundamentally different interpretation. Although he admitted what he called a structural role for religion—that there is a relationship between the various spirit manifestations of God and the points of division in a political system, since people who live together will worship together—he asserted that God stands outside the society altogether and that his various manifestations in society are only refractions of the external God. This is the idea of the one and the many, so brilliantly treated by Johnson (1942). For Robertson Smith, the notion of God was the notion of social unity, which grew as political organization developed; God stands at the apex of a hierarchy of social groupings. For Evans-Pritchard, God does not grow out of the social structure and is not limited by it. Robertson Smith and Evans-

Pritchard are not simply two scholars with disparate opinions on a limited, specific problem. They represent two ways of looking at human life between which there is no possibility of compromise: Evans-Pritchard attaches basic importance to what men think, while Robertson Smith attaches similar importance to what men do, and each scholar observes and weighs the evidence accordingly.

### Kinship and marriage

The second major work of Robertson Smith's that is of importance to modern sociology is *Kinship and Marriage in Early Arabia* (1885). Much of this work is marred—indeed is rendered virtually profitless in parts—by his preoccupation with the stages in the evolution of marriage and with an attempt to demonstrate that kinship was reckoned in the female line before precedence came to be given to male descent. The influence of his friend McLennan on his thinking in this field of study was an unmitigated misfortune. It led Robertson Smith to indulge in wild speculations, which he tried to support by guessing many of the facts. However, when he used authentic data, he did make a number of astute points.

Early Arab society was divided into political groups based on genealogical descent lines, reckoned through males. Robertson Smith did not concern himself with arid verification or refutation of the validity of genealogies. Instead, he was quick to grasp the significance of the consistency between the branching of genealogical lines and the political divisions of tribes. This consistency meant, he said, that the genealogies were manipulated to suit the reality of political divisions and that this was done by altering the length of descent lines to give groups of equal political status equally long descent lines ([1885] 1903, p. 11). Further, he showed some appreciation of the political processes (later to be known as fission and fusion) whereby tribal segments came together to form sections in opposition to like sections, only to dissolve into their separate parts when the situation provoking the combination ceased. His way of putting it was that several groups combined to form a larger whole, "resolving again into their elements" (p. 2). It is to his credit also that in his examination of the processes of alliance and opposition he noted that the occurrence of female names was indicative of greater cohesion in societies where descent was in the male line and polygamy was permitted: female names served to distinguish between paternal half-brothers and thus provided an additional divisive element in

simple patrilineal descent. After offering this interesting idea, however, Robertson Smith abandoned it and used the occurrence of female names as "proof" of an earlier system of kinship through women.

The crux of the problem of tribal organization, as Robertson Smith saw it, lay in the composition of the smallest political groups, out of which the larger ones were built (p. 3). Essentially, these were local communities whose constitution rested on the herding necessities of groups of people moving about, together with their animals, and controlling their own natural resources in water and pasture (p. 41). Robertson Smith conceived of each of these groups as a kindred and, while he rightly saw that consanguineous agnation was relevant within this group but not in larger ones, he also seized the important point that the smallest group was not merely "the family grown large" (p. 5)—that it was a political group and therefore quite unlike a family.

Idiomatic expression in ancient times gave the impression (and still does among Bedouin Arabs) that small political groups were composed of members of a kind of big family. Robertson Smith perceived that terms which appear to mean blood connection may refer to actual consanguineous relationships, or they may be used in the sociological sense of conceptualizing a unity. Thus when phrases such as "one blood" were used to refer to a group or when it was said after the loss of a member through homicide that "our blood has been spilt," it is erroneous to infer that all members of the group were consanguineously related (p. 46). Robertson Smith urged that discussions of kinship be divested of sentimentality and that categories of kinship be regarded not in terms of physiological connections but rather in the more meaningful terms of the duties and obligations which they symbolize. A father, in this sense, is the person who carries the obligations of protector, guardian, and provider of food. In ancient Arabia (and among the present-day Bedouins also) these obligations were acquired by virtue of the payment of bridewealth, as agreed to prior to the marriage contract. This transfer of wealth gave a man procreative rights to his spouse—"what he purchases is the right to have children by her and to have these children belong to his own kin" (p. 130). And further, he acquired the right to delegate the act of procreation to another man without losing his legal relationship to the children born to his wife—hence the saying, "the son is reckoned to the bed on which he is born" (p. 139). If consanguinity did not matter in this close relation-



ship, then obviously it was largely irrelevant when more distant relationships were involved. Kinship, in other words, was a matter of social relationships.

If relationships, then, are to be understood in terms other than those of consanguinity, it is necessary, Robertson Smith suggested, to look at the rights attached to each category. Thus, it would be profitable to regard the family as a property-owning unit, and the delimitation of this particular "kinship unit" would be the range of heirs (pp. 162 ff.). In analyzing marriage, his stress was similar: "Marital rights are rights of property which can be inherited . . ." (p. 105). Leach, writing three quarters of a century later, created something of a stir when he began his analysis of a Singhalese village with the assertion: "Kinship as we meet it in this book is not 'a thing in itself'" (1961, p. 11). He ended his analysis with the same statement, devoting the intervening pages largely to showing that kinship is nothing "except in relation to land and property" (p. 305). (This reductionist argument is discussed in Peters 1963. The composition of local groups among Bedouins, along lines not unlike those suggested by Robertson Smith in his book on kinship, is analyzed in Peters 1965.)

The tribal group of men that Robertson Smith conceived as a kindred was in fact a corporation. Equal responsibility was the privilege of all its members, regardless of the particulars of real or fictive kinship relationships. A group was "a single life" ([1885] 1903, p. 46) and indivisible. It possessed a unique name, leaving no doubt about a person's identity (p. 26), and it had rules of recruitment (pp. 40 ff.). Each corporation had a homeland with its own natural resources (p. 41). It also possessed the reproductive facilities of women (pp. 132-145). Control over its mobile property was achieved by disinheriting women (p. 117) who, had they been permitted to transmit property, could, by marrying outside the group, alienate its property. The corporation had authority to compel its members to accept stated obligations and to outlaw them in case of default (pp. 25, 27). In conclusion, the corporate group, to be effective, required members to be coresidents; since the most impelling obligation of membership was participation in offense and defense, it was necessary to be able to rally members in one place at short notice. Thus, corporations were local communities with a high concentration of agnates.

Later inquiries into the nature of tribal systems in ancient Arabia led Robertson Smith increasingly to ascribe importance to the institution of the

blood feud. There are references to the blood feud in his published works, but the notes he left were too fragmentary to form the basis of a complete analysis; his premature death robbed us of what would surely have been a brilliant work.

E. L. PETERS

[For the historical context of Robertson Smith's work, see the biographies of DURKHEIM; FUSTEL DE COULANGES; for discussion of the subsequent development of his ideas, see MARRIAGE, article on MARRIAGE ALLIANCE; MYTH AND SYMBOL; RELIGION; RITUAL.]

#### WORKS BY SMITH

- (1868) 1912 *Prophecy and Personality: A Fragment*. Pages 97-108 in *Lectures & Essays of William Robertson Smith*. Edited by J. S. Black and George Chrystal. London: Black.
- 1875 *Bible*. Volume 3, pages 634-648 in *Encyclopaedia Britannica*. 9th ed. Edinburgh: Black.
- (1885) 1903 *Kinship and Marriage in Early Arabia*. New ed. London: Black. → See especially the Preface by E. L. Peters to the paperback edition published in 1967 by Beacon.
- 1886 *Sacrifice*. Volume 21, pages 132-138 in *Encyclopaedia Britannica*. 9th ed. Edinburgh: Black.
- (1889) 1927 *Lectures on the Religion of the Semites: The Fundamental Institutions*. 3d ed. New York: Macmillan. → A paperback edition was published in 1956 by Meridian as *Religion of the Semites: The Fundamental Institutions*.
- Lectures & Essays of William Robertson Smith*. Edited by J. S. Black and George Chrystal. London: Black, 1912.

#### SUPPLEMENTARY BIBLIOGRAPHY

- BLACK, J. S.; and CHRYSAL, GEORGE 1912 *The Life of William Robertson Smith*. London: Black. → Contains a bibliography of Robertson Smith's writings.
- BRYSON, GLADYS 1945 *Man and Society: The Scottish Inquiry of the Eighteenth Century*. Princeton Univ. Press.
- CHELHOD, JOSEPH 1955 *Le sacrifice chez les Arabes: Recherches sur l'évolution, la nature et la fonction des rites sacrificiels en Arabie occidentale*. Paris: Presses Universitaires de France.
- DURKHEIM, ÉMILE (1912) 1954 *The Elementary Forms of the Religious Life*. London: Allen & Unwin; New York: Macmillan. → First published as *Les formes élémentaires de la vie religieuse, le système totémique en Australie*. A paperback edition was published in 1961 by Collier.
- EVANS-PRITCHARD, E. E. 1954 *The Meaning of Sacrifice Among the Nuer*. *Journal of the Royal Anthropological Institute of Great Britain and Ireland* 84:21-33.
- EVANS-PRITCHARD, E. E. (1956) 1962 *Nuer Religion*. Oxford: Clarendon.
- FORTES, MEYER 1959 *Oedipus and Job in West African Religion*. Cambridge Univ. Press.
- FRANKFORT, HENRI 1948 *Kingship and the Gods: A Study of Ancient Near Eastern Religion as the Integration of Society and Nature*. Univ. of Chicago Press.
- FUSTEL DE COULANGES, NUMA DENIS (1864) 1956 *The Ancient City: A Study on the Religion, Laws, and*

- Institutions of Greece and Rome*. Garden City, N.Y.: Doubleday. → First published in French.
- GOODE, WILLIAM J. (1951) 1955 *Religion Among the Primitives*. Glencoe, Ill.: Free Press. → A paperback edition was published in 1964.
- HUBERT, HENRI; and MAUSS, MARCEL (1899) 1964 *Sacrifice: Its Nature and Function*. Univ. of Chicago Press. → First published as "Essai sur la nature et la fonction du sacrifice" in Volume 2 of *Année sociologique*.
- JOHNSON, AUBREY R. 1942 *The One and the Many in the Israelite Conception of God*. Cardiff: Univ. of Wales.
- LEACH, EDMUND R. 1961 *Pul Eliya, a Village in Ceylon: A Study of Land Tenure and Kinship*. Cambridge Univ. Press.
- MAUSS, MARCEL (1925) 1954 *The Gift: Forms and Functions of Exchange in Archaic Societies*. Glencoe, Ill.: Free Press. → First published as *Essai sur le don: Forme et raison de l'échange dans les sociétés archaïques*.
- PETERS, E. L. 1960 The Proliferation of Segments in the Lineage of the Bedouin in Cyrenaica. *Journal of the Royal Anthropological Institute of Great Britain and Ireland* 90:29-53.
- PETERS, E. L. 1963 Aspects of Rank and Status Among Muslims in a Lebanese Village. Pages 159-200 in Julian A. Pitt-Rivers (editor), *Mediterranean Countries: Essays in the Social Anthropology of the Mediterranean*. Paris: Mouton.
- PETERS, E. L. 1965 Aspects of the Family Among the Bedouin of Cyrenaica. Pages 121-146 in *Comparative Family Systems*. Edited by Meyer F. Nimkoff. Boston: Houghton Mifflin.
- PETERS, E. L. 1967 Some Structural Aspects of the Feud Among the Bedouin of Cyrenaica. Unpublished manuscript.
- RADCLIFFE-BROWN, A. R. (1952) 1961 *Structure and Function in Primitive Society: Essays and Addresses*. London: Cohen & West; Glencoe, Ill.: Free Press. → A paperback edition was published in 1965 by the Free Press.

## SMOKING

Tobacco was introduced to Europe by Columbus upon his return from the New World. As the habit of smoking spread throughout Europe, Russia, Africa, and the Orient, it met with formidable opposition from rulers, who instigated periodic attacks upon it. In 1604, King James I of England issued a proclamation called "A Counterblast to Tobacco." He and others, however, were unsuccessful in eradicating the use of tobacco, which soon became an export of considerable economic importance to the new American colonies. In the United States tobacco remains economically important; it contributes to national income, and its yield of tax revenue plays a significant role in the budget of the federal government.

Although social and religious controversies occasionally have flared between those segments of society which found smoking pleasurable and those which believed it to be a foul-smelling, unnatural, and loathsome habit, scientists showed little interest in the effects of smoking for some five centuries after its introduction to Europe.

In 1900 vital statisticians reported an increase in the incidence of cancer of the lung. Beginning in 1930, scientists from various parts of the world have reported vital statistics suggesting a relationship between certain populations' increasing consumption of cigarettes and an increased incidence of lung cancer, heart disease, and other illnesses in the same populations. Since the mid-1950s these results have become sufficiently convincing to cause the following health agencies to declare that smoking is a significant health hazard: the British Medical Research Council; the cancer societies of Denmark, Norway, Sweden, Finland, and the Netherlands; the American Cancer Society; the American Heart Association; the Joint Tuberculosis Council of Great Britain; and the Canadian National Department of Health and Welfare. However, cigarette manufacturers have steadfastly maintained that the evidence linking cigarettes and disease is not conclusive.

In any study of smoking, an initial problem arises with regard to the definition of who is a cigarette smoker and, even more problematic, who is an ex-smoker, light smoker, medium smoker, or heavy smoker. There is no precise definition as to what are the actual behavioral (motions), physiological, or psychological referents of smoking. For example, do all individuals who "smoke heavily" inhale to the same extent? One individual might light 40 cigarettes a day but puff only one-eighth of each, whereas another individual, claiming to smoke only 12 cigarettes a day, could inhale each of these down to the last quarter inch. We also do not know the relationship between reported smoking (through verbal self-report or responses to a questionnaire) and actual smoking. Data relevant to these preliminary questions would contribute greatly to the precision of subsequent research on smoking.

Within the past decade social scientists have become interested in the smoker, as evidenced by the appearance of numerous surveys and other research reports on the psychosocial characteristics of smokers. The vast majority of these have focused only on cigarette smokers, and the findings reported below will relate primarily to them.

The reader interested in a detailed review of



this subject may consult five reviews that include comprehensive bibliographies of the original literature. Two of the reviews appeared in 1960—one in the United States (Matarazzo & Saslow 1960), and one in Scotland (Kissen 1960). A third review was published by Lawton (1962) and a fourth by Horn (1963). The most recent review (Hochbaum 1964) was prepared as part of the highly publicized report of the advisory committee to the surgeon general of the United States Public Health Service.

Several individual studies are of especial significance because of their scope, sampling, or over-all design. Haenszel, Shimkin, and Miller (1956) surveyed the smoking habits of a cross section (40,000) of the total population of American men and women 18 years of age and older. Sackrin and Conover (1957) followed this by further study of 18,000 of the same 40,000 individuals. Kallner (1958) studied the entire population of Israel. Lilienfeld (1959) studied 4,400 adults in Buffalo, New York. Horn and his associates (1959) polled the total population of high school students (22,000) in Portland, Oregon, while Salber and her associates (see Salber & MacMahon 1961; Salber et al. 1962; Salber, Welsh, & Taylor 1963; Salber & Worcester 1964; Salber et al. 1961; 1963) have published a series of related studies of 6,810 high school students in Newton, Massachusetts. In England, Bothwell (1959) conducted a study of the smoking habits of 8,314 Oxfordshire schoolchildren, while Chave and Schilling and their associates (1959) studied another group of 3,500 London youngsters. Also, Eysenck and his co-workers (1960) attempted to measure some personality characteristics of a stratified sample of 2,360 smoking and nonsmoking adult British males and followed this with a more detailed study of 3,194 additional males (Eysenck 1963).

**Demographic variables.** Although social scientists have only recently begun to study the cigarette smoker, there already seem to be a number of variables which are consistently found to be associated with smoking. These variables are reviewed below, although, because of space limitations, no attempt has been made to review the findings of every study on each variable. Rather, the more significant or better-established findings of a particular study have been presented, and the original articles can be referred to for more detail. The reader interested in the attempts of health educators and social scientists to understand why people smoke and in some of the methods currently being explored in the United States to help smokers give up their habit should consult the report by Mausner and Platt (1966).

**Age, sex, and marital status.** With respect to age, data from both sides of the Atlantic are similar and indicate that regular smoking usually begins in mid-adolescence to late adolescence, although a small percentage of youngsters start cigarette smoking earlier. Boys are inclined to smoke somewhat earlier and more heavily than girls (Bothwell 1959; Cartwright et al. 1959; Haenszel et al. 1956; Hochbaum 1964; Horn et al. 1959; Horn 1963; Kissen 1960, p. 368; Matarazzo & Saslow 1960; and Salber et al. 1961). Among women, the incidence of smoking is increasing. In their study in Newton, Massachusetts, Salber and Worcester (1964) report that the number of women, particularly Jewish women, who smoke may soon exceed the number of men who smoke.

Among both sexes and all ages, there is a greater percentage of smoking among divorced and widowed individuals than among those either married or never married. Although this has been established for the United States, comparable data for other countries are not available.

**Income, occupation, and education.** The five reviews suggest that yearly income is not related in any consistent manner to smoking. However, socioeconomic class, when defined only in terms of occupational and educational level, does appear to bear a consistent relationship to smoking; individuals at the lowest occupational levels start to smoke earlier and in greater numbers than those in all other groups. For example, white-collar groups (professional workers, managers, and so forth) typically contain fewer smokers than are found among craftsmen, foremen, salespersons, and similar groups. Possibly it is because some individuals in the latter groups have higher yearly incomes than do some white-collar workers that the relationship between smoking and socioeconomic class, more broadly defined to include income and place of residence as well, is not clearly established.

Apparently there is no clear linear relationship between the highest educational level attained by maturity and smoking (see especially Lilienfeld 1959). However, among high school students those with highest achievement and those in college preparatory courses are considerably less likely to smoke than those in business and technical courses. An interesting twenty-year longitudinal study of a group of Harvard undergraduates indicates that individuals who major in the arts and letters, education, and social sciences later become smokers in greater numbers than do students whose career choice is in the physical sciences (Heath 1958; McArthur et al. 1958).

*Urban-rural differences.* In all studies examining urban-rural differences, rural farm populations, including both sexes and all ages, were found to contain a smaller percentage of smokers than either the rural nonfarm or the city populations. Rural nonfarm persons closely resemble urban dwellers in their smoking habits. Women and girls in large cities in the northeastern United States (e.g., New York City and Boston) smoke more than do women and girls in cities in other areas of the United States. It is unknown whether a similar pattern prevails when cosmopolitan centers such as London, Paris, and Rome are used as a basis for comparison.

*Race and religion.* In the English-speaking countries which have been studied, there appears to be no relationship between race and smoking, the proportion of smokers being approximately equal between the Caucasian and Negro subsamples.

It is well known that devout followers of some religions do not smoke at all (e.g., Hindus, Muslims, Mormons). Studies of young men and women of high school and college age in the United States (Horn et al. 1959; Salber & Worcester 1964; Straits & Sechrest 1963) have reported that significantly more followers of the Roman Catholic and Jewish faiths were found to smoke than Protestants in these age groups. In the Newton, Massachusetts, study, Salber and Worcester (1964, p. 36) found that Catholic men and boys smoked more than other males, while Jewish women and girls smoked more than other females.

*Psychological variables.* In addition to sociological and demographic variables, some psychological variables have also been explored.

*Intelligence and achievement.* Present available evidence suggests that there is no direct relationship between intelligence, as measured by tests of IQ, and smoking (Matarazzo & Saslow 1960; Hochbaum 1964).

The same is not true for academic achievement during adolescence, however. The previously mentioned studies in Portland, Oregon; Buffalo, New York; Newton, Massachusetts; and a study in London, England (Davis, as reported in Kissen 1960, p. 369) indicate that those students who are academically inferior to their age mates or classmates have a significantly greater tendency to be smokers. This relationship between achievement and smoking, of course, does not establish that one of the two factors is a cause of the other. It can be stated, however, on the basis of consistent evidence, that groups of young men and women who are academically less successful than their

peers contain a higher proportion of smokers than is found among their more successful classmates. Also, smokers are found to take the scholastically less demanding academic programs (vocational preparation in contrast to a college preparatory curriculum). Smokers date more but engage in fewer extracurricular activities, including sports of some types, than do nonsmokers. No interpretation of these disparate facts has yet achieved wide acceptance.

*Personality.* Studies attempting to relate personality variables to smoking suffer from the serious handicap that there exists today no highly valid, universally accepted measure either of global personality or of any specific personality dimension. (This assertion quite probably will find support among serious students of personality.) One result is that seldom do any two studies, even those allegedly measuring the same trait (e.g., neuroticism), utilize the same methods of assessment; therefore they cannot be compared meaningfully.

Matarazzo and Saslow (1960), Kissen (1960), Lawton (1962), Horn (1963), and Hochbaum (1964) have presented reviews of the few published studies in this area. The results, meager and poorly supported as they are, suggest the following. In most English-speaking countries roughly half of the population are smokers. Studies using stratified as well as nonstratified samples (some small, some large) reveal the presence of a slightly higher number of "extraverts" and "neurotic," "anxious," and "tense" individuals among the smokers as compared to the nonsmokers. However, although the differences in the mean scores of smokers and nonsmokers are statistically significant, they typically reflect a difference in response to only one or two questionnaire items. Two of the most extensive (3,194 individual subjects) and best-designed studies on this subject (Eysenck et al. 1960; Eysenck 1963) utilized stratified samples of individuals in the British Isles. In the second one Eysenck (1963, pp. 114-115) concludes: "Degree and type of smoking are . . . related to extraversion, in the sense that pipe smokers are introverted, as are non-smokers; light, medium and heavy smokers are extraverted, increasingly so in that order." The degree of extraversion among ex-smokers is found to be indistinguishable from that among light smokers. Such a strong conclusion may not be justified, however, since the mean extraversion scores, on the 31-item extraversion measure used by Eysenck, were as follows for the groups studied (in the order in which they are mentioned above): 17.07, 17.63, 17.60, 18.70, 18.95, and 17.71. Although several of the groups



differed significantly from each other because of the large number of subjects, the mean difference between the highest and lowest groups was only 1.88 items on the 31-item personality measure. Similar small but statistically significant differences between smokers and nonsmokers have been found on the "anxiety" scale used by Matarazzo and Saslow (1960, p. 499) and on the "tension," "psychosomatic," and "neuroticism" measures used by other workers.

Referring to the research described above in their 1960 review, Matarazzo and Saslow concluded

. . . while smokers do differ from non-smokers in a variety of characteristics, none of the studies has shown a single variable which is found exclusively in one group and is completely absent in the other . . . [this] is especially true for the variables measuring personality characteristics. . . . Examination of the means, standard deviations, ranges, percentages, etc. . . . makes clear that while group trends suggest the smoker to be more "neurotic," *on the average*, there are still many individual smokers with neuroticism, or anxiety . . . scores lower than those of many non-smokers. . . . Thus, a clear-cut smoker's personality has not emerged from the results so far published in the literature. . . . It is hard to believe that they [the half of a country's population who smoke] would share in common one personality "type." This is not to imply, however, that the various psychological dimensions along which smokers have been shown, as a group, to differ from non-smokers may not suggest an important single process, or processes, underlying these various demonstrated (but small) differences. Further research may indeed so systematize the disparate findings. (1960, pp. 508-510)

Thus, a number of studies report small differences in mean scores, suggesting that compared to the nonsmoker the average smoker has a slight tendency to be more extraverted, outgoing, adventurous, tense, anxious, inclined to drink more alcohol and coffee, date earlier, watch more television, see more movies, drive an automobile earlier, read fewer books per month, play fewer sports, and belong to fewer clubs and school organizations. However, a large number of individual smokers may show few or none of these characteristics. In addition, no unifying theme, trait, or personality dimension has been abstracted from these separate characteristics. For the present it seems best to conclude that any psychological dynamics that may differentiate the smoker from the nonsmoker have not yet been isolated.

**Initiation of smoking.** Phanishayi (1951), in India, was one of the first to study variables associated with the initiation of smoking. His research

was followed by that of McArthur and his associates (1958) with Harvard College undergraduates in Boston; Bothwell (1959) in Oxfordshire, England; Chave, Schilling and their associates (1959) in London; Horn and his associates (1959) in Portland, Oregon; Cartwright et al. (1959) in Edinburgh, Scotland; Morison and Medovy (1961) in Winnipeg, Canada; and Salber and MacMahon (1961) in Newton, Massachusetts.

These studies consistently have identified parental smoking as one of the most important predisposing factors in smoking among school-age children. As mentioned above, most smokers appear to have begun smoking between the ages of 10 and 18. If both parents smoke, the probability that their children will begin to smoke is several times that of children with nonsmoking parents. When only one parent smokes, the incidence of smoking among the offspring falls midway between that of the other two groups. Published data also suggest a higher frequency of smoking among children with older siblings who smoke.

The relationship of some other sociopersonal factors to initiation of the smoking habit is less clear-cut. In general, the studies suggest that youngsters' beginning to smoke is related to: (a) curiosity about smoking; (b) conformity pressures among adolescents; (c) need for status among peers, including self-perceived failure to achieve peer-group status or satisfaction; (d) the need for self-assurance; and (e) striving for adult status (see the reviews by Hochbaum 1964; Horn 1963). However, it is difficult to measure the strength of such needs, as well as their relative influence, and therefore these relationships should be considered tentative.

There is no convincing evidence that beginning to smoke in childhood or young adulthood is a sign of rebellion against parents or other authority figures, although such a hypothesis has been suggested by several writers.

Most writers have suggested that once the smoking habit has been established the factors associated with its continuation very likely are different from those associated with its initiation. Although the evidence is far from conclusive, it appears that situational tension-reduction is an important motivational factor in perpetuating the smoking habit. Also, a small group of smokers attest that they continue to smoke because it is pleasurable.

Many smokers report considerable dissatisfaction with smoking but, despite repeated attempts, have found themselves unable to break the habit. In fact, Horn (1963, p. 364), who sampled the adult popu-

lation of the entire United States, reports that only 14 per cent of regular cigarette smokers state that they consider the habit pleasurable, safe, and worth the cost.

**Discontinuation of smoking.** Why do smokers not give up the habit that so many state they would like to shed? Despite pronouncements from health authorities in all parts of the world, the proportion of smokers reporting successful discontinuation of smoking is from 10 to 20 per cent in males and 3 to 10 per cent in females (Haenszel et al. 1956, p. 24; Hammond & Percy 1958, p. 2,956; Cartwright et al. 1959, p. 726; and Horn 1963, p. 391).

Immediately following the 1964 report of the surgeon general of the United States, 8.6 per cent of the smoking students at two U.S. colleges reported that they stopped smoking as a result of the health hazard identified in this report (Katahn et al. 1964). Again, a greater percentage of male smokers (10.3 per cent) than female smokers (5.9 per cent) reported quitting. In addition, more natural science majors (18.2 per cent) quit smoking than did liberal arts majors (6.7 per cent). The same students took a story-completion test requiring them to create both an imaginary story and an estimate of the amount of time which in reality would have transpired during the occurrence of the sequence of events in the story. Ex-smokers tended to respond with a longer time perspective than did students who continued to smoke, and the authors suggest that these individuals might be more likely to worry about long-term health hazards.

Immediately before and following the surgeon general's report, several researchers attempted to find effective means for discouraging initiation and encouraging cessation of the smoking habit. Horn (1960) used a variety of educational approaches with high school students in Portland, Oregon, but he reported only minimal success in preventing the initiation of smoking. Both Horn (1964, personal communication) and Lawton (1964, personal communication) have attempted to help adult smokers discontinue smoking through the use of weekly educational meetings (Horn in Washington) and weekly "group therapy" sessions (Lawton in Philadelphia). With both techniques, success rates were not much higher than the 10 to 20 per cent "spontaneous" successes reported by other writers. The Horn and Lawton studies raise an important question for all cessation studies: who is an ex-smoker? Is he an individual who quits for a week, a month, a year, or forever? It is both a common observation and an established statistical fact that

many ex-smokers resume the habit after varying lengths of time.

**Comment.** After reviewing the above literature, the present authors have concluded that, in all probability, people begin to smoke for essentially the same reasons that they adopt other habits, such as eating certain foods, using certain brands of soap, wearing lipstick, going to see the Beatles, and so forth. That is, initiation of smoking appears to be a result of the millions of dollars spent yearly by (cigarette) manufacturers to influence people to do so; the standards and fads extant in youthful and adult social circles; the examples provided by parents and families; the depth of commitment to personal religious codes; and, finally, some psychological and personality characteristics (e.g., extraversion, tension, neuroticism, etc.) of the individual himself. In our opinion, the last category (personality characteristics) is the least influential in determining whether a given individual will begin to smoke, whereas the other factors (especially the amount spent on world-wide advertising) have considerable influence. Once started, however, continuation or discontinuation appears to be, in large part, a function of the sociopsychological characteristics of the individual.

JOSEPH D. MATARAZZO AND RUTH G. MATARAZZO

[Directly related are the entries DRINKING AND ALCOHOLISM; DRUGS.]

#### BIBLIOGRAPHY

- BOTHWELL, P. W. 1959 The Epidemiology of Cigarette Smoking in Rural School Children. *Medical Officer* 102:125-132.
- CARTWRIGHT, ANN; MARTIN, F. M.; and THOMSON, J. G. 1959 Distribution and Development of Smoking Habits. *Lancet* (1959) no. 2:725-727.
- CHAVE, S. P. W.; and SCHILLING, R. S. F. et al. 1959 The Smoking Habits of School Children. *British Journal of Preventive and Social Medicine* 13:1-4.
- EYSENCK, H. J. 1963 Smoking, Personality and Psychosomatic Disorders. *Journal of Psychosomatic Research* 7:107-130.
- EYSENCK, H. J. et al. 1960 Smoking and Personality. *British Medical Journal* [1960], no. 5184:1456-1460.
- From Epidemiology to Ecology—Smoking and Health in Transition. 1966 *American Journal of Public Health* 56, no. 12.
- HAENZSEL, WILLIAM; SHIMKIN, MICHAEL B.; and MILLER, HERMAN P. 1956 *Tobacco Smoking Patterns in the United States*. U.S. Public Health Monograph No. 45; Public Health Service Publication No. 463. Washington: Government Printing Office.
- HAMMOND, E. CUYLER; and PERCY, CONSTANCE 1958 Ex-smokers. *New York State Journal of Medicine* 58:2956-2959.
- HEATH, CLARK W. 1958 Differences Between Smokers and Non-smokers. *AMA Archives of Internal Medicine* 101:377-388.



- [HOCHBAUM, G. M.] 1964 Psycho-Social Aspects of Smoking. Pages 359-379 in U.S. Surgeon General's Advisory Committee on Smoking and Health, *Smoking and Health: Report of the Advisory Committee to the Surgeon General of the Public Health Service*. Public Health Service Publication No. 1103. Washington: Government Printing Office.
- HORN, DANIEL 1960 Modifying Smoking Habits in High School Students. *Children* 7:63-65.
- HORN, DANIEL 1963 Behavioral Aspects of Cigarette Smoking. *Journal of Chronic Diseases* 16:383-395.
- HORN, DANIEL et al. 1959 Cigarette Smoking Among High School Students. *American Journal of Public Health* 49:1497-1511.
- KALLNER, GERTRUDE 1958 Smoking Habits of the Population. Unpublished manuscript, Israel, Central Bureau of Statistics.
- KATAHN, MARTIN; SPIELBERGER, CHARLES D.; and JUDSON, ABE J. 1964 Smokers and Ex-smokers: Reactions of College Students to the Surgeon General's Cancer Report. *Psychonomic Science* 1:323-324.
- KISSEN, DAVID M. 1960 Psycho-Social Factors in Cigarette Smoking Motivation: A Review. *Medical Officer* 104 365-372.
- LAWTON, M. P. 1962 Psychosocial Aspects of Cigarette Smoking. *Journal of Health and Human Behavior* 3: 163-170.
- LILIENFELD, ABRAHAM M. 1959 Emotional and Other Selected Characteristics of Cigarette Smokers and Non-smokers as Related to Epidemiological Studies of Lung Cancer and Other Diseases. *Journal of the National Cancer Institute* 22:259-282.
- MCARTHUR, CHARLES; WALDRAN, ELLEN; and DICKINSON, JOHN 1958 The Psychology of Smoking. *Journal of Abnormal and Social Psychology* 56:267-275.
- MATARAZZO, JOSEPH D.; and SASLOW, GEORGE 1960 Psychological and Related Characteristics of Smokers and Nonsmokers. *Psychological Bulletin* 57:493-513.
- MAUSNER, BERNARD; and PLATT, ELLEN S. 1966 Behavioral Aspects of Smoking: A Conference Report. *Health Education Monographs Supplement*, No. 2. → Presents the highlights of a three-day conference held at Beaver College (Glenside, Pennsylvania) attended by 39 scientists and health educators.
- MORISON, JAMES B.; and MEDOVY, H. 1961 Smoking Habits of Winnipeg School Children. *Canadian Medical Association Journal* 84:1006-1012.
- PHANISHAYI, R. A. 1951 Causes of Smoking Habit in College Students: An Investigation. *Journal of Education and Psychology* 9:29-37.
- SACKRIN, SEYMOUR M.; and CONOVER, ARTHUR G. 1957 *Tobacco Smoking in the United States in Relation to Income*. U.S. Department of Agriculture, Market Research Report, No. 189. Washington: The Department.
- SALBER, EVA J.; and MACMAHON, BRIAN 1961 Cigarette Smoking Among High School Students Related to Social Class and Parental Smoking Habits. *American Journal of Public Health* 51:1780-1789.
- SALBER, EVA J.; MACMAHON, BRIAN; and WELSH, BARBARA 1962 Smoking Habits of High School Students Related to Intelligence and Achievement. *Pediatrics* 29: 780-787.
- SALBER, EVA J.; WELSH, BARBARA; and TAYLOR, S. V. 1963 Reasons for Smoking Given by Secondary School Children. *Journal of Health and Human Behavior* 4: 118-129.
- SALBER, EVA J.; and WORCESTER, JANE 1964 Change in Women's Smoking Patterns. *Cancer* 17:32-36.
- SALBER, EVA J. et al. 1961 Smoking Habits of High School Students in Newton, Massachusetts. *New England Journal of Medicine* 265:969-974.
- SALBER, EVA J. et al. 1963 Smoking Behavior, Recreational Activities and Attitudes Toward Smoking Among Newton Secondary School Children. *Pediatrics* 32: 911-918.
- STRAITS, BRUCE C.; and SECHREST, LEE 1963 Further Support of Some Findings About the Characteristics of Smokers and Non-smokers. *Journal of Consulting Psychology* 27:282.
- U. S. SURGEON GENERAL'S ADVISORY COMMITTEE ON SMOKING AND HEALTH 1964 *Smoking and Health: Report of the Advisory Committee to the Surgeon General of the Public Health Service*. Public Health Service Publication No. 1103. Washington: Government Printing Office.

### SNOUCK HURGRONJE, CHRISTIAAN

Christiaan Snouck Hurgronje (1857-1936), Dutch orientalist and colonial official, was educated at Leiden University. He was a lecturer at his alma mater between 1880 and 1889 and for 17 years thereafter was an adviser on religious and native affairs to the Netherlands Indies government. From 1906 until his death, he was professor of Arabic and Islamic institutions at Leiden and concurrently (until 1933) adviser to the Ministry of Colonies.

Snouck Hurgronje is best known as one of Europe's two nineteenth-century pioneers in the scientific study of Islam (the other being the Hungarian Ignaz Goldziher). His rigorously critical, historical approach, based on close scrutiny of scriptural texts and Arabic commentaries, shed new light on such crucial problems as the origins of the faith and the role of Muhammad in its growth and development, and the nature of the *hadith* (the recorded usages and traditions of the Prophet and the early Muslim communities). Snouck was the first to recognize the central role in Islam of the *fiqh*, i.e., a repository of quasi-judicial regulations pertaining not only to legal and social affairs but also to the relation of the individual to Allah, and of the *idjma*, i.e., the "consensus" on interpretations of the *fiqh* reached by the four recognized schools of Islam, which is binding on all Muslims. At the same time Snouck Hurgronje critically re-examined Islam as a historical reality, reconstructing, on the basis of the Qu'ran (Koran) and the *hadith*, the origins of the first Muslim communities, with special emphasis on the adaptations to pre-existing rituals instituted by Muhammad. Snouck was also the first to clarify

the interplay between pre-Islamic usages ('*āda*) and *fiqh* and hence the pattern of diversity within unity. Islamic eschatology, especially that centering on the Mahdi, and Islamic mysticism were other important fields opened up by him, which were subsequently thoroughly explored in the Indonesian setting by some of his outstanding Dutch students.

Although Snouck was recognized as a leading authority on Islam in European and Muslim lands even during his lifetime, his other achievements have as yet received only scanty scholarly attention. Apparently without formal academic training in anthropology, Snouck produced ethnological and ethnographic studies of lasting importance. It was his concern for Islam as a living reality, his sharp sense of observation and his mastery of both scriptural and vernacular languages, and his careful field work that permitted him to do so. His classic two-volume work, *Mekka* (1888–1889), the result of his sojourn there in 1885, typically combined a reconstruction from original texts of the Holy City's history with vivid, detailed descriptions of daily life, gathered in "coffee shops, *ditwans*, mosques and living quarters." Snouck almost intuitively grasped the concept of "culture," as is evidenced in the preface to his standard work on the Achinese of northern Sumatra (1893–1894): "To him who really wants to penetrate the Muhammadan factor in the life of a people, children's games, adult entertainments, profane literature, the organization of a village or province—all these are in many respects as important as are the books used in religious instruction, the mystical orders propagating in the land, the position of the scribes" (1893–1894, vol. 1, p. xii). A similar perceptiveness pervades his observations on Javanese Islam, published (in part only) pseudonymously in 1891–1892 as a series of letters from a retired Javanese district head.

As lecturer in a training institute for colonial officials from 1880 on and as the first Dutchman to gain on-the-spot knowledge of the Indonesian ("Djawa") colony in Mecca, Snouck Hurgronje was throughout his career deeply and often passionately involved in the problems confronting a European colonial power ruling millions of Indonesian Muslims. His studies on the Achinese and Javanese were written while he served the colonial government in an advisory capacity; the insights he gained during his Achinese mission were instrumental in terminating a desultory, decades-old and Muslim-inspired war against the Dutch. But more than a mere adviser, Snouck was in fact the

originator and architect of modern Dutch Islamic policy, a policy basically adhered to until the end of Dutch rule in Indonesia in 1942. Its twin principles were the broadest toleration of religious life, on the one hand, and rigorous repression of Islamic, especially pan-Islamic, political agitation, on the other. Although the dividing line was as difficult to determine in theory as in practice, particularly after the advent of Islamic reformism in Indonesia (after Snouck had returned to Holland), the twentieth-century Islamic renaissance in the archipelago was doubtless facilitated by the colonial government's adherence to Snouck's directives.

However, there is a tragic paradox in the fact that, in spite of his lasting influence on Islamic policies, Snouck Hurgronje was least successful as a colonial statesman in the wider sense of the word. He was one of the very few Dutchmen with a bold vision of the ultimate destiny of peoples under colonial domination and virtually the lone propounder of the associationist ideal that envisaged a future political union between Holland and Indonesia as equal partners sharing a common cultural bond. He was a fervent proponent of rapid Westernization through education, which he believed would readily supplant indigenous, including Muslim, orientations, especially among the Javanese aristocracy. Snouck argued in favor of progressive Indonesianization of the administration, which was to go hand-in-hand with a planned withdrawal of European supervisory personnel. Especially in the years after World War I, he became one of the most resolute Dutch advocates of ultimate Indonesian independence.

Dutch colonial practice in the twentieth century, however, took a course which was diametrically opposite to that which Snouck Hurgronje advocated. His departure from Indonesia in 1906 had been caused by his inability to persuade the colonial authorities to supplement pacification (in Achin) with meaningful social reconstruction. The iron logic, merciless irony, and uncompromising vehemence with which he argued his case (his later, presumably highly critical, advisory opinions to the Ministry of Colonies have not yet been published) made him the idol of many liberals but also the *bête noire* of diehard colonial conservatives. In order to undercut his influence and that of his colleague Cornelis van Vollenhoven in the training of colonial administrators at Leiden University, these conservatives in 1925 established a rival department of Indonesian studies at the University of Utrecht.



Snouck's failure as a colonial administrator should not obscure his lasting stature in the field of Islamic studies. Social scientists have not yet fully come to appreciate his contributions to many of their specialties.

HARRY J. BENDA

[See also ASIAN SOCIETY; COLONIALISM; ISLAM; and the biography of SCHRIEKE.]

#### WORKS BY SNOUCK HURGRONJE

- 1888-1889 *Mekka*. 2 vols. The Hague: Nijhoff. → Volume 1: *Die Stadt und ihre Herren*. Volume 2: *Aus dem heutigen Leben*.
- (1889) 1931 *Mekka in the Latter Part of the 19th Century: Daily Life, Customs and Learning, the Moslems of the East-Indian Archipelago*. Leiden: Brill; London: Luzac. → First published in German. A translation of Volume 2 of Snouck Hurgronje 1888-1889.
- (1891-1892) 1924 *Brieven van een wedono-pensioen*. Pages 111-248 in Christiaan Snouck Hurgronje, *Verspreide geschriften*. Volume 4: *Geschriften betreffende den Islam in Nederlandsch-Indië*. Bonn and Leipzig: Schroeder; Leiden: Brill. → First published in Java in a Dutch newspaper *De Locomotief*, between January 7, 1891, and February 22, 1892.
- (1893-1894) 1906 *The Achehnese*. 2 vols. London: Luzac. → First published as *De Atjehers*. Translation of extract from the 1893-1894 edition was provided by Harry J. Benda.
- 1923-1927 *Verspreide geschriften*. 6 vols. in 7. Bonn and Leipzig: Schroeder; Leiden: Brill. → Volume 1: *Geschriften betreffende den Islam en zijne geschiedenis*. Volume 2: *Geschriften betreffende het mohammedaansche recht*. Volume 3: *Geschriften betreffende Arabië en Turkije*. Volume 4: *Geschriften betreffende den Islam in Nederlandsch-Indië*. 2 vols. Volume 5: *Geschriften betreffende taal- und letterkunde*. Volume 6: *Boekaankondigingen; verscheidenheden; registers. Ambtelijke adviezen van C. Snouck Hurgronje, 1889-1936*. 3 vols. The Hague: Nijhoff, 1957-1965.
- Selected Works of C. Snouck Hurgronje*. Edited by G. H. Bousquet and J. Schacht. Leiden: Brill, 1957.

#### SUPPLEMENTARY BIBLIOGRAPHY

- BENDA, HARRY J. 1958 Christiaan Snouck Hurgronje and the Foundation of Dutch Islamic Policy in Indonesia. *Journal of Modern History* 30:338-347.
- BOUSQUET, GEORGES-HENRI 1938 *La politique musulmane et coloniale des Pays-Bas*. Paris: Hartmann. → A partial translation of this book was published in 1940 as *A French View of the Netherlands Indies*, by Oxford University Press.
- DREWES, GERARDUS W. J. 1957 *Snouck Hurgronje en de Islamwetenschap: Herdenking van de 100e geboortedag van Prof. Dr. C. Snouck Hurgronje, 10 februari 1957*. Leiden: Universitaire Pers.
- PEDERSEN, JOHANNES 1957 *The Scientific Work of Snouck Hurgronje*. Leiden: Brill.
- PIJPER, GUILLAUME F. 1957 *Islam and the Netherlands*. Leiden: Brill.
- PIJPER, GUILLAUME F. 1961 *De islampolitiek der Nederlandse regering*. Pages 209-222 in Henri Baudet and Isaac J. Brugmans (editors), *Balans van beleid. Terugblik op de laatste halve eeuw van Nederlandsch-Indië*. Assen (The Netherlands): Van Gorcum.

VAN NIEL, ROBERT 1957 Christiaan Snouck Hurgronje: In Memory of the Centennial of His Birth. *Journal of Asian Studies* 16:591-594.

## SOCIAL ACCOUNTING

See NATIONAL INCOME AND PRODUCT ACCOUNTS; NATIONAL WEALTH; for related material, see also INCOME DISTRIBUTION; INPUT-OUTPUT ANALYSIS.

## SOCIAL ANTHROPOLOGY

See under ANTHROPOLOGY.

## SOCIAL BEHAVIOR, ANIMAL

- |                                           |                     |
|-------------------------------------------|---------------------|
| I. THE FIELD                              | J. P. Scott         |
| II. PRIMATE BEHAVIOR                      | Irven DeVore        |
| III. THE REGULATION OF ANIMAL POPULATIONS | V. C. Wynne-Edwards |

### I THE FIELD

Social behavior may be defined as any behavior which is stimulated by or has an effect upon another animal of the same species. So defined, almost all behavior may have some degree of sociality—and it is therefore proper to speak of behavior as highly social or only minimally so—but there is very little behavior which can be called truly asocial. Social behavior is therefore very nearly equivalent to behavior in general, as the same behavior may be social or not, depending upon its source of stimulation and its object. To qualify the above definition further, there are occasions, particularly in cases of domestication or hybridization, in which social behavior can be elicited by a member of another species.

**Behavior patterns.** A behavior pattern may be defined as a segment of behavior having a definite adaptive function, and therefore, as a basic functional unit of behavior. For example, a chicken approaching a pan of water, lowering its bill into the water, and then raising its head is exhibiting a pattern of behavior which has the function of providing the animal with water.

**Distribution of social behavior.** The paramoecium, so commonly studied in elementary biology courses, is a highly developed protozoan which exhibits a variety of behavior patterns. One of these is a simple form of sexual behavior in which the paramoecia fasten themselves together and afterward exchange nuclei and divide. Another pattern occurs under unfavorable environmental conditions. When a drop of water containing several paramoecia is allowed to dry slowly, the animals will huddle together and thus protect each

other against drying out. Such behavior has the function of shelter or comfort seeking. In addition, this one-celled animal shows other patterns of behavior which are only incidentally social. One of these is an investigatory or exploratory pattern of behavior in which the animal drives itself ahead with its cilia in a spiral path and thus comes in contact with various objects. Depending on the nature of the objects, it may remain in contact or back off and start out in another direction. Finally, its pattern of ingestive behavior is extremely simple, consisting of driving water laden with food particles toward its gullet by means of cilia, and it involves only starting or stopping the beating cilia.

More complex patterns of behavior having basically the same functions are present in the lower invertebrates, but only in the arthropods and vertebrates can we find new functions. Male crayfish, for example, will fight with each other during the breeding season, an example of *agonistic* behavior. Certain insects show highly developed patterns of behavior for the care of the young, thus exhibiting care-giving, or *epimeletic*, behavior. The young larvae, by their movements, may signal for care and attention, exhibiting care-soliciting, or *et-epimeletic*, behavior. Likewise, ingestive behavior assumes a highly social function, especially among ants, which feed each other in a relationship that Wheeler (1923) called *trophallaxis*.

All of these behavioral functions are highly developed in vertebrates. In addition, superior eyesight permits vertebrates to develop patterns of social behavior which involve mutual imitation; this *allelomimetic* behavior is commonly seen in schools of fish, flocks of birds, and herds of mammals, where each animal follows the movements of those next to him, resulting in highly coordinated group movements (Scott 1958).

**Systems of behavior.** As behavior becomes more complex in the animal kingdom, a given species develops several alternate patterns of behavior for each general function. A paramecium, for example, has only one simple pattern of ingestive behavior and no patterns of agonistic behavior. In contrast, a mouse attacked by another mouse may fight back, run away, assume a defensive posture, or become limp and passive. Each of these patterns of behavior has the function of adapting to the same stimulus, and the attacked mouse usually goes through all or part of this behavior repertory until one response is finally selected on the basis of its effect. Such a group of behavior patterns having a common general function, and usually a common physiological basis, is a *behavioral sys-*

*tem*. Behavioral systems are quite similar to physiological systems, and many of them are directly related, as for example, sexual behavior and the reproductive system. Like physiological systems, behavioral systems do not function entirely independently of each other, and in various species they may be organized on a somewhat different basis. In the dog and other members of the family Canidae, sexual and eliminative behavior are closely related to each other, whereas in birds the two are unrelated.

The nature of behavior patterns and their organization into a system is largely a result of the special heredity of the species concerned. A primary task in studying the social behavior of any species is to describe its special behavior patterns and their organization into behavioral systems. This behavioral inventory, or ethogram, as Tinbergen has named it (1951), provides the essential raw material for understanding social organization or conducting experimental work.

**Development of behavioral systems.** In most of the higher animals, behavioral systems are incompletely developed at birth or hatching. In young mammals the ingestive system is present in an immature form, the newborn animal taking its food only by nursing. The development of behavior patterns for ingesting solid food comes later. Some systems of behavior may be completely absent, and all of them tend to be loosely organized together as *et-epimeletic* behavior. As the animal grows older, both behavior patterns and behavioral systems become differentiated from each other, partly as the result of learning and partly because of the processes of growth and maturation.

### Social organization

**Social relationships.** A social relationship may be defined as regular and predictable behavior between two individuals, usually of the same species. In some animals these relationships may be organized largely as the result of heredity, but in higher animals relationships develop as a combined result of hereditary factors and learning. When two chickens meet for the first time, each presents a problem of adaptation to the other, and one solution is to attack. The result is a fight, and usually one chicken wins and the other loses. When they meet again, there is a tendency to repeat the behavior, so that in the long run one individual forms a strong habit of winning and the other of losing. Since one of the principles of learning is the "law of effect" (which might also be called the law of least effort), the winning chicken reduces its attacks as time goes on, and the whole



relationship is reduced to one of threat and avoidance behavior rather than actual fighting. This is an example of a typical dominance-subordination relationship. [See LEARNING, *articles on INSTRUMENTAL LEARNING and REINFORCEMENT.*]

Hereditary factors also help to determine the development of the relationship. If one bird is a male and the other a female, the male almost invariably wins. Among females, large hens are likely to win over small ones, and aggressive breeds over peaceful ones (Guhl 1953).

Thus, the development of a social relationship may be considered as a process of differentiation of behavior, affected by both heredity and learning. As will be seen below, the development of social relationships in young animals is a slower and more complex process.

Social relationships can be classified in two different ways. In Carpenter's system, classification is based on the different kinds of individuals present in the system (1934). In howling monkeys and most other mammals there are three major kinds of individuals (males, females, and young) combined in six kinds of social relationships (see Table 1). The same scheme applies to all social vertebrates, but it becomes more complex in some of the social invertebrates. In bees there are three kinds of adults—males (drones), females (queens), and sterile females (workers)—as well as the young. In addition, the workers exhibit different behavior at different stages of development. In termites there may be as many as three different kinds of individuals capable of reproduction and as many more different kinds of workers.

Table 1 — Social relationships classified according to age and sex

	Male	Female	Young
Male	Male-male		
Female	Male-female	Female-female	
Young	Male-young	Female-young	Young-young

Source: After Carpenter 1934.

Social relationships may also be classified according to the behavioral system or systems involved. On this basis a large number of social relationships are theoretically possible, but actually observed, important social relationships are relatively few in number. There are four which involve one behavior system only. *Dominance-subordination* relationships are developed in any animal system where agonistic behavior is found. *Sexual* relationships depend of course on sex behavior. *Leader-follower* relationships are based on allelomimetic behavior and result from an unequal

degree of mutual imitation. Finally, *mutual-care* relationships, often seen in the mutual grooming of primates, are formed when both animals exhibit epimeletic behavior.

There are two important relationships composed of combined systems: trophallaxis and care-dependency. Trophallaxis involves mutual feeding and mutual care and is very important in the social insects. The workers feed the young and in turn lick off secretions from them. In mammals, birds, and other vertebrates which care for their young, epimeletic and et-epimeletic behavior are combined to produce the relationship of care-dependency.

Both of these schemes of classification may be combined for systematic study. For example, the behavior exhibited in a male-male relationship may be examined in detail and classified as dominance-subordination, leader-follower, or whatever it may be in the particular animal society.

The number of possible social relationships which an animal develops depends on the size of the group. Carpenter has presented a simple formula for the total number:  $n(n-1)/2$ , which is of course the number of possible combinations between any two individuals in a group of  $n$  individuals. When examined in more detail, this formula has certain interesting properties which have implications for the social organization of groups. If we consider groups of different sizes, starting with one and going up to ten, we see that an individual added to a group adds a number of relationships equal to the number of individuals already present, i.e., an individual joining a group of three adds only three new relationships, but adds ten to a group of ten. There is a very rapid increase in the number of relationships as the group gets larger, resulting in extremely complex organization. Experiments indicate that hens in a flock of one hundred can recognize every other individual, making a theoretical total of 4,950 relationships in the flock. Social activity tends to increase in proportion to the number of social relationships.

**Primary socialization.** In highly developed societies the individuals must become related to particular members of the species. An ant colony would become totally disorganized if its members reacted toward members of other colonies in the same way as to their own. Ants would wander from one nest to another, overcrowding one and leaving young unprotected in others. Therefore, at some period in their development, all individuals must learn to differentiate their own family or colony from others. As Fielde (1903) showed,

there is a critical period in early development in which the immature ant can be transferred from one nest to another or even to that of another species and subsequently act as if it belonged with the new individuals. This transference of social relationships occurs regularly in the slave-raiding ants, a species of red ant commonly found in North America. The raiders remove the pupae from nests of black ants and take them to their own, where the black individuals subsequently rear and care for the offspring of the raiders.

A similar phenomenon takes place in all highly social animals which have been studied so far. Birds hatched in incubators or removed from their nests in early stages of development and reared by hand become attached to the human species and often fail to react to their own. In precocial birds, such as chickens, ducks, geese, and turkeys, the formation of the first social relationship takes place within the first 24 hours after birth. In the more slowly developing altricial birds, such as song sparrows, jackdaws, or doves, the process takes place at a later period in development, and the degree of attachment to the human handler varies inversely with the age at which the young are taken from the nest [see IMPRINTING].

Among mammals, also, the process varies with speed of development. In sheep and other herd animals the young lamb becomes strongly attached to its mother within the first few days, whereas in the more slowly developing dog the period of socialization begins at about three weeks of age, reaches a peak at about six or seven weeks, and slowly declines thereafter. The process of socialization is of course a reciprocal one, the parents becoming attached to their offspring as well. In sheep or goats the mother becomes attached to her offspring within the first two hours or so and thereafter will repulse any strange infant. Primary socialization has been little studied in primates other than man, but it takes place within the first six months in rhesus monkeys, as also appears to be the case in human infants [see INFANCY, article on THE EFFECTS OF EARLY EXPERIENCE].

The time at which primary socialization takes place is obviously a critical period in development, determining which individuals shall be the close relatives of the young animal. Later he may exhibit appropriate patterns of sexual behavior and other types of social interaction toward similar individuals [see AFFECTION; SOCIALIZATION, article on PSYCHOLOGICAL ASPECTS].

What is the nature of the process involved in primary socialization? In the young animal there are positive behavioral mechanisms such as the

"following" reaction of a young chick or lamb, which brings it into contact with members of its own species. There are also negative mechanisms, particularly the behavior patterns of escape, which develop somewhat later and prevent socialization with other individuals by inhibiting contact. As to the nature of the process itself, in precocial birds it obviously takes place in the absence of food rewards, and, under experimental conditions, without any obvious external rewards or reinforcement of any kind. While food rewards do affect the behavior of a young mammal, experiments have shown that the process of socialization takes place without them. All evidence indicates that the process is an internal one and is closely connected with emotional responses. Experimental studies indicate that any kind of strong emotion, whether it be that of loneliness, fear, pain, or hunger, will speed up the process of forming a social bond. While the capacity to form such a bond is strongly developed in the infant animal, the same process can take place at any time in later life at a somewhat slower rate, except when it is completely suppressed by the development of interfering behavior patterns (Scott 1962).

### Animal and human societies

The kind of social organization developed by a species depends upon the nature of its behavior patterns and the ways in which they are organized into behavioral systems. Thus we may state a general law: Social behavior is a determinant of social organization.

**Aggregations.** The simplest kind of animal society is the temporary aggregation. Such groups occur widely in the animal kingdom and were extensively studied by W. C. Allee, who was able to show in a large number of cases that living in groups promotes the survival of individuals, whether they are protozoans or vertebrates (Allee et al. 1949). This kind of temporary aggregation is based on shelter- or comfort-seeking behavior.

A second kind of temporary aggregation is based on sexual behavior. This may involve only two individuals, as it does in paramecia, or vast numbers, as in the mating swarms of the palolo worm of the Pacific. The great majority of invertebrate societies belong to one or the other of these temporary aggregations.

In contrast, the higher invertebrates, and many vertebrates, form long-lasting groups with highly complex internal organization. Insect groups first received scientific attention, but it is now realized that the social behavior of insects is somewhat specialized, being limited by certain anatomical



peculiarities such as the external skeleton and the relatively inefficient compound eyes. Because of these limitations, insects are small, relatively short-lived, and unable to respond to complex visual stimuli at a distance.

**Insect societies.** The most highly developed insect societies belong to two orders, the Hymenoptera, including ants, bees, and wasps, and the Isoptera, or termites. In both orders the societies consist of permanent groups associated with nests that are constructed by the combined efforts of the members, and the most prominent social behavior is care giving. In an ant colony two workers meet, feel each other with their antennae, and if one has recently fed, it will regurgitate a drop of honeydew to be eaten by the other. This, together with care of young, constitutes most of their social behavior and is what Wheeler termed trophallaxis (1923).

In any given species of insect, social organization tends to be rigid and stereotyped, repeating itself generation after generation. There is relatively little modification by experience, and the combined efforts of the individuals produce a stable social environment. Alfred Emerson points out that insect societies have many of the characteristics of a single organism, and he has called them "supra-organisms" (see in Allee et al. 1949).

The peak of development of the insect society is reached by the common honeybee, which maintains a permanent social group. In contrast to ants, in which each new colony is formed by a mated pair, large numbers of old bees accompany a queen moving to a new hive. Mating takes place at this time, and the fertile female, or queen, continues to reproduce for a long period, sometimes two or three years. The vast majority of the colony members are short-lived sterile females or workers which go through a regular series of activities as they pass through a few weeks of active life, first working around the hive and finally going out to forage for food. As von Frisch discovered, foraging bees are able to communicate to other workers the direction and distance at which food is found by means of body movements, or "dances," when they return to the hive (1950). Present evidence indicates that the nature of the signals and the ability to understand them are determined by heredity rather than learning.

**Vertebrate societies.** The societies of vertebrate animals are also related to basic anatomy and physical capacities. Vertebrates have an internal skeleton and hence are capable of continuous growth. They can become immensely large, and their life spans may be enormously extended in

comparison with insects. The capacity for learning is consequently much more valuable and plays a more important role than it does among invertebrate societies, so that there are many examples of cultural as well as biological inheritance. Furthermore, the vertebrate eye is vastly more efficient than the compound eye of arthropods, making mutual imitation or allelomimetic behavior possible, particularly for animals active in the daytime.

There is an enormous variety among vertebrate societies, ranging from temporary aggregations, like those of the lower invertebrates, to societies far exceeding those of the insects in complexity. The most highly developed societies are found among animals which occupy dominant ecological positions.

In the water, fish are still the dominant form of life, and their societies have taken two forms. One of these, the school, which is based on allelomimetic behavior, is very widely found. In herring or mackerel the fish move in groups throughout their lives, even spawning in a mass and paying no attention to the eggs or young. In contrast, there are many kinds of fish, like the sunfish and stickleback, which emphasize nest building and the care of the young. In these cases the male guards the nest and sometimes the young for a short period after hatching. The period of dependency remains short, however, as in most cases the young must feed themselves. The males guard definite boundaries around the nest and thus exhibit territoriality (Tinbergen 1953).

Amphibians, which usually occupy a relatively inferior ecological niche and are under constant pressure for survival, show almost no social organization except temporary mating aggregations and a few examples of care of the young. The same can be said for many reptiles, although some lizards develop dominance and territoriality and alligators build nests and briefly protect their young.

Birds, which are the dominant form of life in the air, show a great variety of highly developed societies. One of the characteristics of bird societies is the phenomenon of seasonal change. For example, during their brief mating season, song sparrows develop one sort of social organization involving sexual behavior, territoriality, and care of the young and then spend the rest of the year in flocks which are much like schools of fish. Since the young are produced as eggs which have to be incubated, a high degree of coordination of care-giving behavior and physiological processes is necessary, and in many birds this is exhibited by both

sexes, which combine to build the nest and care for the young. Many species feed their young, and pigeons and other members of the dove family nourish their offspring with crop milk. A "nuclear family" of male, female, and young is characteristic of these and many other bird societies, but in chickens and other members of the same family, mating is polygamous, and the brood of young is raised by the female alone.

In mammals there is an enormous variety of societies differing in structure and degree of complexity. Among rodents, woodchucks are relatively solitary animals, reacting to each other with agonistic behavior except during the brief mating and reproductive seasons (Bronson 1964). At the opposite extreme are the prairie dogs, which run to groups numbering thousands of individuals. Each prairie dog colony is subdivided into territories, usually occupied by a sexually mature male, a group of females, and their young. Each year the adults in the territory move out to new locations on the edge of the colony, leaving the burrows behind for the next generation. Thus the oldest and most experienced members of the colony are found on its edges, where adaptation is more difficult (King 1955).

Other examples of mammalian societies are the great herds of hoofed mammals. In a typical society, such as that of mountain sheep, the two sexes mingle during the mating season. Males compete in individual combat, and the more successful ones do most of the mating. During the rest of the year there are separate male and female herds, the latter being the group in which the young are raised. Among these herds a definite leadership organization develops, the older animals of either sex tending to lead in their respective groups.

Some carnivores are quite solitary in their habits, but others, like wolves and dogs among the Canidae and lions in the cat family, live in permanent groups. These are, however, much smaller than those of the herd animals on which they prey.

Thus, we see that there is an enormous number of different kinds of animal societies. There is great variation among human societies, but the variety possible in other animals far exceeds it. It is therefore of considerable interest to see what kinds of social organization are present in man's closest biological relatives, the other members of the Primate order.

**Primate societies.** Like other important orders of mammals, the primate group includes a wide variety of species, ranging from the small "bush

babies," or galagos, of Africa to the large manlike apes and man himself. Primates can be at least as different from one another as cats are from bears in the order Carnivora. Their chief limitation of variation is an ecological one. Except for man, they live entirely in tropical and subtropical regions, and while many primates like to play with water, none of them is truly aquatic.

That primates other than man are highly social was first established by Carpenter's classical study of the howling monkey (1934). These are tree-dwelling primates of Central America. A band of males, females, and young wanders from tree to tree, eating fruits and leaves and rarely descending to the ground. Their most obvious kind of social behavior is allelomimetic, since they are constantly following each other. The more active males generally take the lead, but there is no single established leader in their wanderings. Agonistic behavior is greatly reduced and expressed almost entirely as the vocalization that gives the species its name. Females in estrus pass from male to male, and there is no indication of sexual jealousy or rivalry over females. The most important social relationship within the group is that between mother and offspring, the mother constantly carrying and caring for her baby for the first two years of its life. Care-giving behavior is also shown by the males if a young animal falls to the ground or is threatened in some way. Different bands have roughly marked territories, and groups keep each other apart by means of threats.

Like man, baboons are a plains-living primate, and for this reason their social life has particular interest. Biologically they have evolved toward a doglike structure, often running on all fours and developing enormous teeth. However, they are largely vegetarian rather than hunting animals, often eating grass which they pick and eat a blade at a time. They readily adapt to living off farm crops, particularly melons and vegetables, and can be enormously destructive (Washburn & DeVore 1961; DeVore 1965).

As with howlers, a large part of their social behavior is allelomimetic, and they move in a group of males, females, and young which may include thirty to one hundred individuals. Agonistic behavior is more prominent than in howlers, and the males develop a definite dominance order. As the group moves, the most dominant male stays in the center, while the more subordinate animals keep their distance from him and each other. Living on the plains, the group is constantly threatened by large predators such as lions and cheetahs. Females and their young tend to group themselves toward



the center of the group, near the dominant male, this being the safest spot. When a member of the group is attacked or threatened, all the males combine in its defense. Consequently, baboons are seldom actually harmed by predators. When in estrus, a female may mate with the most dominant male first but then move on to others as he becomes satiated. As with howlers, there is no evidence of sexual jealousy or permanent bonds between particular males and females. The most important social relationship is that between mother and offspring; in fact, all adults are extremely responsive to young animals and will approach and care for them if the mother permits. As the infant becomes more independent, it joins a play group of other young animals within the group. In general, the baboon society gives the impression of a closely knit and highly cooperative group.

Macaques are related to baboons, and their behavior is somewhat similar. The Indian macaque, or rhesus monkey, commonly lives near human habitations and is often fed by people. Carpenter studied an artificial colony of rhesus monkeys on Santiago Island off Puerto Rico and demonstrated that a dominance order between the males was an important part of their social life (1942). As with baboons, females in estrus pass from male to male. Overt aggressive behavior is perhaps more common than in baboons, possibly because of the disorganizing effect of human contact. Southwick found that where rhesus troops went into cities and competed for food, there were many cases of serious injury (1963).

The behavior of the Japanese macaque is quite similar. Imanishi reports that the animals spend their lives within a troop composed of males, females, and young (1960). The most dominant male usually occupies the central position, surrounded by females, while the younger and less dominant males keep to the outside of the troop and also precede it when it is moving. As with baboons, there is no subdivision of the troop into families or any permanent consortship between individual males and females.

The Indian langur, or Hanuman monkey, is a tree-living species and exhibits much less aggressive behavior than the macaques. However, Jay has described a definite dominance order among males (1965a; 1965b). All females are subordinate to all males, except when an infant is threatened, and their own dominance order is not a clearly marked one. Sexual behavior is relatively unimportant, as the females have only a five-day estrus period per month and this is absent during pregnancy and lactation. Females have some tendency to consort

with the most dominant male, but there is no absolute correlation, and in one group studied by Jay the most favored male was a low-ranking one. Male-female consortships never last more than a few hours. In contrast to baboons, males pay little attention to the young, but infants, and particularly the newborn ones, are the object of much attention from all females in the troop. Males tend to take the lead when the group moves.

The manlike apes have been very difficult to study in the wild. In his original study Nissen (1931) was able to see very little besides the retreating chimpanzees disappearing into the tops of the tropical forest. More recently, Kortlandt located a favorable area on a Congo plantation where chimpanzees were protected and were in the habit of coming into the open (1962). Reynolds and Reynolds were able to keep track of a group of chimpanzees in the Budongo forest of Uganda with the aid of a large group of spotters scattered through the dense vegetation (1965). Unlike baboons, chimpanzees do not live in compact groups, but forty to fifty individuals may occupy an area of six to eight square miles. Except for females and their offspring, there are no permanent associations between individuals, and temporary groups of males and females, females and young, males, and groups containing all kinds of individuals are found in approximately equal numbers. Agonistic behavior is uncommon, and there are few indications of dominance and subordination. During estrus the same female may mate with several males in succession, and no fighting is observed between them. Chimpanzees appear to keep in touch with each other by vocalization, and on many occasions groups will join in vocalization over a large area. There is little evidence of leadership under most conditions and no reports of defense of territorial boundaries. Thus, chimpanzees form a very loosely knit but mutually tolerant society, with loose temporary associations between individuals, except for the mother-offspring relationship.

Goodall has studied chimpanzees in the comparatively open forests of the Gombe Stream Reserve in Tanganyika and has been able to follow the behavior of identifiable individuals over a period of several years (1965). Her more detailed accounts of behavior confirm the general picture of chimpanzee social organization. In addition, she has observed occasional predatory behavior and meat eating. These chimpanzees also exhibit use of tools, in particular, poking twigs or rolled-up leaves down the holes of termites in order to collect them for food.

The behavior of gorillas under natural conditions is also surprising. With their immense size these animals have little to fear from predators and within their own social groups are placid and peaceable. They are strict vegetarians, and the two sexes are so much alike that they are almost impossible to distinguish at any distance. Their sex organs are very small, and sex behavior plays a very minor part in their social life. Unlike chimpanzees, gorillas live in small compact groups containing both sexes and all ages of individuals. According to Schaller, there is a strong system of leadership, all members of the group responding to one of the older males (1963). Dominance is expressed almost entirely in the form of allowing precedence, and it is strongly related to age.

The only one of the manlike apes that shows anything like the human nuclear family of male, female, and their offspring is the gibbon. These long-armed tree-dwelling animals of southeast Asia are highly aggressive in both sexes, and the largest social group is a male, a female, and their immature offspring. Other groups are kept away by threats and fighting. Rather than supposing this to be the beginning of evolution toward a nuclear family, we can advance the hypothesis that the gibbon group is formed by a reduction of the troop to the smallest unit which is still biologically functional.

Primate societies do not present an evolutionary history of human social organization. Rather, each species has evolved its own social organization, producing an astonishing variety against which it is possible to compare and contrast human social organization. The one thing which is constant in all primate societies so far studied is the emphasis on the care of the young, particularly by the mother but also by unrelated males and females in the highly social forms. Agonistic behavior varies a great deal, as does sexual behavior. Most primates show regular periods of estrus, although there is some indication of a tendency toward a more extended receptivity in female chimpanzees. Was the social organization of our primitive human ancestors more like troops of rhesus and baboons, or was there a tendency to form temporary separate groups based on sex and age, as in the chimpanzee? The almost universal tendency toward division of labor in human societies suggests the latter. The animal data also suggest that the human nuclear family is perhaps not the beginning but the end point of human social organization, resulting from situations which demand great fluidity; that is, the nuclear family represents the ultimate reduction of family organization to the simplest

unit which is still capable of all the functions of biological reproduction [see FAMILY].

**The evolution of social organization.** Social organization is based on behavior, and behavior leaves no fossils. Therefore, any reconstruction of the evolution of animal societies must always be hypothetical and based on what we know about living animals. Assuming that the more lowly organized forms of living animals are similar to those which existed millions of years ago, we can suppose that there are two general bases for the initiation of social life. One of these is shelter-seeking behavior, in which animals stay together because the bodies of their fellows form a favorable environment. If this sort of behavior is extended, it should result in the animals' following each other around, i.e., *allelomimetic* behavior. Animals that do the latter must have both motor equipment for rapid coordination and the necessary sense organs to keep in touch with each other. This kind of social group approaches its highest development in schools of fish, flocks of birds, and herds of mammals.

The other kind of social behavior which we may assume to be primitive is sexual behavior, which results in more efficient means of reproduction than scattering the sex cells broadcast. Retaining the eggs after they are fertilized and giving them care after they are laid still further increases the efficiency of reproduction, and the height of this kind of behavior is seen in two widely different parts of the animal kingdom—societies of insects and those of many birds and mammals.

These two basic kinds of social organization may evolve independently of each other, or both may occur together in the same species. In many cases social organization is quite independent of physical form, and closely related species may be either highly social or relatively unsocial. For example, the deer family range from the highly social elk through the more solitary Virginia deer to the moose, which is almost completely solitary except during the brief mating season and the association of a calf with its mother. Among rodents there is the prairie dog, a ground squirrel that lives in colonies of thousands of individuals, and at the opposite extreme, woodchucks, which lead highly solitary lives except for the brief necessities of mating and rearing the young. Even in the social insects there are close relatives of honeybees which build solitary nests. Thus there is no consistent relationship between biological form and social behavior, and the evolution of societies has become largely, though not entirely, independent of the anatomy of the individual. Consequently, in search-



ing for the origin of human sociality, it is not possible to find anywhere the recapitulation of human social prehistory. The study of animal societies can at most suggest ideas and provide a basis of comparison.

**The problem of basic human nature.** In addition to biological evolution, human societies have added a new dimension: cultural evolution. The beginnings of this can be seen among the higher animals where the offspring learn fears from the previous generation. However, human language has enormously increased the potentiality of passing information along from one generation to another, and written language permits its accumulation over almost infinite times and in enormous amounts.

Compared to biological evolution, the study of cultural evolution is still in its infancy. Very little progress was made in the theory of biological evolution until the mechanism of heredity was discovered. Reasoning by analogy, we may suggest that the development of an adequate theory of cultural evolution will in part depend on our knowledge of the mechanism of cultural heredity, namely learning, and its application to cultural phenomena.

Unlike biological evolution, which moves in terms of generations, cultural change can occur with extreme rapidity and can be measured in terms of years. This raises the possibility that cultural change may move so fast that it exceeds the biological capacities of man in respect to social behavior. [See CULTURE, article on CULTURE CHANGE.]

Here again we are handicapped in our scientific information. There is no way of restoring precultural man to life and discovering what sort of social existence he led. Judging from existing animal societies, however, we may assume that his social behavior was well adapted to a particular kind of stable social organization. Beyond this, all we can do is to observe present-day human beings and study their social behavior under varying conditions. We know that sexual behavior plays an important role. Instead of definitely limited periods of sexual behavior, human females have very nearly reached the situation of constant receptivity, if not constant estrus, and there is no seasonal limitation on either sex. We know also that, because of the relatively slow developmental rate of human infants, care-giving behavior is highly important, as is the care-soliciting behavior of the young. It is likewise obvious that care-giving behavior is often extended to adults as well as to children and that it exists in both sexes. Young children are there-

fore one of the focal points of human societies, and as with all primate societies, the relationship between old and young is an important one. As with any highly social species, we may assume that human beings have evolved behavioral capacities for the control of agonistic behavior, partly through forming habits of not fighting while young and partly through the formation of dominance-subordination relationships. These facts are almost self-evident. It is not as well recognized that human beings are also highly allelomimetic, with strong tendencies to do what others around them are doing. Because of differences in age and sex, allelomimetic behavior may be difficult in mixed groups, and we can observe a tendency for human groups to sort themselves out on the basis of age and sex. [See AGGRESSION; IMITATION; LEADERSHIP.]

Besides these major types of advanced social behavior, human beings continue to show all of the more primitive sorts of social behavior, for example, ingestive, eliminative, shelter-seeking, and investigative. In regard to the last, another characteristic of human beings is that they are highly curious. We can conclude that human beings show all of the basic kinds of social behavior in strongly developed form, and we can advance the hypothesis that any successful form of cultural organization of society must provide for reasonably satisfactory development and expression of these types of behavior. Conversely, any form of organization which attempts to completely suppress or distort a fundamental kind of behavior will result either in the dissolution of the society or maladaptive behavior on the part of the individual within it.

J. P. SCOTT

[Directly related are the entries ETHOLOGY; IMPRINTING; INSTINCT; PSYCHOLOGY, article on COMPARATIVE PSYCHOLOGY. Other relevant material may be found in AFFECTION; AGGRESSION, article on PSYCHOLOGICAL ASPECTS; COLLECTIVE BEHAVIOR; COMMUNICATION, ANIMAL; CULTURE; EVOLUTION; GENETICS; GROUPS; IMITATION; KINESICS; SEXUAL BEHAVIOR, article on ANIMAL SEXUAL BEHAVIOR; SOCIAL PSYCHOLOGY.]

#### BIBLIOGRAPHY

- ALLEE, W. C. et al. 1949 *Principles of Animal Ecology*. Philadelphia: Saunders.
- BRONSON, F. H. 1964 Agonistic Behaviour in Woodchucks. *Animal Behaviour* 12, no. 4:470-478.
- CARPENTER, CLARENCE R. 1934 A Field Study of the Behavior and Social Relations of Howling Monkeys (*Alouatta palliata*). *Comparative Psychology Monographs* 10, no. 2; serial no. 48.
- CARPENTER, CLARENCE R. 1942 Sexual Behavior of Free Ranging Rhesus Monkeys (*Macaca mulata*). 2 parts.

II

PRIMATE BEHAVIOR

*Journal of Comparative Psychology* 33:113-162. →  
I: Specimens, Procedures and Behavioral Characteristics of Estrus. II: Periodicity of Estrus, Homosexual, Autoerotic and Non-conformist Behavior.

- CLARK, WILFRED E. LEGRIS 1960 *The Antecedents of Man: An Introduction to the Evolution of the Primates*. Chicago: Quadrangle.
- DEVORE, IRVEN (editor) 1965 *Primate Behavior: Field Studies of Monkeys and Apes*. New York: Holt.
- FIELDE, ADELE M. 1903 Artificial Mixed Nests of Ants. *Biological Bulletin* 5:320-325.
- FRISCH, KARL VON 1950 *Bees: Their Vision, Chemical Senses, and Language*. Ithaca, N.Y.: Cornell Univ. Press.
- GOODALL, JANE 1965 Chimpanzees of the Gombe Stream Reserve. Pages 425-473 in Irvén DeVore (editor), *Primate Behavior: Field Studies of Monkeys and Apes*. New York: Holt.
- GUHL, A. M. 1953 *Social Behavior of the Domestic Fowl*. Technical Bulletin No. 73. Manhattan: Kansas State College, Agricultural Experiment Station.
- IMANISHI, KINJI 1960 Social Organization of Subhuman Primates in Their Natural Habitat. *Current Anthropology* 1:393-407.
- JAY, PHYLLIS 1965a The Common Langur of North India. Pages 197-249 in Irvén DeVore (editor), *Primate Behavior: Field Studies of Monkeys and Apes*. New York: Holt.
- JAY, PHYLLIS 1965b Field Studies. Volume 2, pages 525-587 in Allan M. Schrier, Harry F. Harlow, and Fred Stollnitz (editors), *Behavior of Nonhuman Primates*. New York: Academic Press.
- KING, JOHN A. 1955 *Social Behavior, Social Organization, and Population Dynamics in a Black-tailed Prairie Dog Town in the Black Hills of South Dakota*. Contributions from the Laboratory of Vertebrate Biology, No. 67. Ann Arbor: Univ. of Michigan.
- KORTLANDT, ADRIAAN 1962 Chimpanzees in the Wild. *Scientific American* 206, no. 5:128-138.
- NISSSEN, HENRY W. 1931 A Field Study of the Chimpanzee. *Comparative Psychology Monographs* 8, no. 1; serial no. 36.
- REYNOLDS, VERNON; and REYNOLDS, FRANCES 1965 Chimpanzees of the Budongo Forest. Pages 368-424 in Irvén DeVore (editor), *Primate Behavior: Field Studies of Monkeys and Apes*. New York: Holt.
- SCHALLER, GEORGE B. 1963 *The Mountain Gorilla: Ecology and Behavior*. Univ. of Chicago Press.
- SCOTT, J. P. 1958 *Animal Behavior*. Univ. of Chicago Press.
- SCOTT, J. P. 1962 Critical Periods in Behavioral Development. *Science* 138:949-958.
- SOUTHWICK, CHARLES H. (editor) 1963 *Primate Social Behavior; an Enduring Problem: Selected Readings*. Princeton, N.J.: Van Nostrand.
- TINBERGEN, NIKOLAAS 1951 *The Study of Instinct*. Oxford: Clarendon.
- TINBERGEN, NIKOLAAS 1953 *Social Behaviour in Animals, With Special Reference to Vertebrates*. London: Methuen.
- WASHBURN, SHERWOOD L.; and DEVORE, IRVEN 1961 Social Behavior of Baboons and Early Man. Pages 91-105 in Sherwood L. Washburn (editor), *Social Life of Early Man*. Viking Publication in Anthropology, No. 31. Chicago: Aldine.
- WHEELER, WILLIAM M. 1923 *Social Life Among the Insects*. New York: Harcourt.

Man is a large, bipedal, diurnal primate closely related to the living great apes. Human behavior is based on a rich social heritage made possible by a tool-dependent culture and the unique properties of language. However, in most of the fundamental features of his social life, such as prolonged care of immature offspring and lifelong association between related adults, man is a typical Old World primate. Because man is most closely related to the Old World monkeys and apes, the following discussion is confined primarily to these forms and directed to those issues of most interest to the social sciences.

**Living primates.** There are about six hundred varieties of living nonhuman primates, divided into some fifty genera and two hundred species. Because of this extraordinary diversity, ranging from such forms as the small, insectivorelike tree shrews and "mouse lemurs" to the great apes and man, it is sometimes assumed that the behavior of living primates can be arranged along an evolutionary scale of increasing behavioral similarity to man. However, no living primate is the ancestor of any other, and many varieties have had separate evolutionary histories for tens of millions of years. Every living species has survived by specialized adaptations, and while it is possible to make some broad generalizations about the physical characteristics of prosimians, monkeys, and apes (Clark 1960), behavioral comparisons are more difficult. Washburn and Hamburg (1965) have recently discussed the classification of the primate order from a behavioral point of view.

The many varieties of monkeys alone range in size from creatures weighing less than a pound to others weighing more than one hundred pounds. They have exploited a wide variety of jungle and open woodland habitats, frequently achieving population densities of one hundred or more individuals per square mile. The Old World monkeys (Cercopithecidae) are divided into two subfamilies, the Cercopithecinae and the Colobinae; the latter group is distinguished by a specialized stomach capable of digesting large quantities of mature leaves. Brief observations have been made of several species of Colobinae, but the only form studied in detail is the common langur of India and Ceylon, *Presbytis spp.* (Jay 1965). Among the Cercopithecinae, numerous studies have been made of the two closely related ground-adapted forms, the African baboon (*Papio*) and the Asian macaques (*Macaca*). Of the other two basic groups of



Cercopithecinae, the mangabeys (*Cercocebus*) and the guenons, or vervets (*Cercopithecus*), a long-term field study has been made of only one, a ground-adapted species, *Cercopithecus aethiops* (Struhsaker 1965).

The four apes, the Pongidae, are the gibbon (including the siamang), the orangutan, the gorilla, and the chimpanzee. All of the apes differ from the monkeys in that the trunk is "short, wide and shallow . . . the lumbar region is short . . . and the shoulder and arm muscles are especially adapted to abduction, flexion, and rotation" (Washburn & Hamburg 1965, p. 9). Because of this adaptation to climbing and the ability to hang or swing from branches, to brachiate, apes have a "vertical" posture, thus differing from the horizontal, quadrupedal monkeys, and it is no surprise that in their structure and bodily movements the apes closely resemble man. Physically, and to some extent behaviorally, man is most like the African great apes—the gorilla and the chimpanzee. But apes are confined to heavily wooded areas, and some of the behavioral patterns which have enabled man to travel long distances in open country are more closely paralleled in the behavior of terrestrial monkeys, such as baboons and macaques. In fundamental senses such as hearing, eyesight, and smell and in the organization of the brain, the viscera, and the reproductive organs man, the apes, and the Old World monkeys share a basic biological pattern which distinguishes them from all other mammals, including New World monkeys and other primates.

**Field studies of primates.** Although attempts were made to study the African apes soon after their discovery in the nineteenth century, most reports of monkey and ape behavior in the wild remained anecdotal until Carpenter observed the howling monkeys of Panama in 1931. In succeeding years, other field studies were attempted, but these were usually brief; then, in the 1950s, investigators in Japan, the United States, and Europe independently initiated a variety of field studies emphasizing long-term, systematic observation of primates living in natural habitats.

Most field studies have been made on monkey species that spend large amounts of time on the ground, where observation conditions are easiest. Because of their similarity to man, all the living apes except the orangutan have been the subject of at least one major field study. Long-term studies of free-ranging groups in monkey colonies have also contributed valuable data. The Japan Monkey Center has kept records of individual monkeys in its colonies since the early 1950s. Beginning with

Altmann in 1958 (see Buettner-Janusch et al. 1962), various investigators have restudied the rhesus monkey colony established on Cayo Santiago by Carpenter in 1938; in fact, since 1958 almost all the members of this free-ranging colony have been tattooed for identification. Because the life cycle of primates is so long, field workers cannot ordinarily ascertain sibling or mother-offspring bonds between adults, and yet these relationships are proving to be very important in some species. (For a more complete history of field studies see Southwick 1963, pp. 1–6; and Washburn et al. 1965.)

Laboratory studies of primates, especially of the common Indian rhesus monkey, have continued without interruption since the 1930s. Contemporary views of such topics as discrimination, operant conditioning, perception, and learning have been summarized by Schrier, Harlow, and Stollnitz (1965). Recent studies of hemoglobins, blood serums, chromosomes, the nervous system, and the skin have been reported by Buettner-Janusch (1963–1964). There are two journals devoted entirely to primate studies, *Primates* and *Folia primatologica*, and there are many films of primate behavior.

The formerly facile generalizations about monkey and ape behavior are no longer tenable. Although only about a dozen of the several hundred varieties of living primates have been studied to date, it is already clear that primates exploit a wide variety of ecological niches, that they live in different kinds of social groups, and that different species display quite different temperaments. Some species are distributed over large geographical areas, and group structure and behavior have been found to vary significantly between different populations of the same species (Jay 1968). The unique life history of every individual produces distinctive patterns of temperament and social behavior among the members of a single group. All of these variables make it inevitable that there will be important exceptions to many of the following generalizations.

### Primate group structure

Most vertebrate social groups undergo a variety of changes in group composition during the year. Changes are correlated with such phenomena as the abundance of food, the mating season, the birth season, and the maturation of offspring. During the mating season adult males may drive off young males and form harems of receptive females; mothers may leave the group altogether when they give birth, living apart until their off-

spring are able to support themselves, or females with immature offspring may form separate bands of their own. Some or all of these groups may participate in seasonal migrations. The result is a "population" with changing aggregations of individuals during the yearly cycle.

The social group of the Old World monkeys and the apes is very different. Membership in the social group is usually continuous, and group composition may remain stable over many years. Typically, an individual is born, matures, leads its adult life, and dies in the same group. While there are interesting exceptions to these generalizations, they are so characteristic of a wide variety of monkey groups that it is convenient to speak of a distinctive group organization, the "troop." A troop so defined is a group of several adults of both sexes, together with juveniles and infants, which maintains a social identity and spatial unity that persists through the annual cycle and transcends the life-span of individual members. Such a troop is easily recognizable to the field observer because its boundaries are typically defended against outsiders. This social organization describes the basic social group of such Old World monkeys as lan-

gurs, baboons, macaques, vervets (*Cercopithecus aethiops*), and colobus; of the New World howling monkey (*Alouatta palliata*), and perhaps spider monkeys (*Ateles geoffroyi*) (see Table 1). Less complete studies suggest that this type of group organization may also be characteristic of at least some species of mangabeys and some other species of *Cercopithecus*. The troop is also characteristic of *Lemur macaco* and probably of some other species of lemur. The description of the troop would apply to the social organization of the mountain gorilla, except that members may move in and out of some mountain gorilla groups rather freely.

Old World monkey and gorilla troops average about 25 individuals, but normal groups may be as small as 10 or number more than 200. The fact that primates normally remain in one social group throughout the year means that group members with needs and motivations as different as a small, helpless infant and a large, potentially dangerous adult male must accommodate each other. Among most vertebrates, when antagonisms arise between adults or between adults and mature offspring, usually during mating or birth periods, the result is a change in group structure. In some pri-

Table 1 — Types of social groups among primates

Type of group	Prosimians	New World monkeys	Old World monkeys	Apes	Preagricultural man
Solitary individuals (usually nocturnal or crepuscular)	Tree shrew "Lesser lemur" Aye-aye Lepilemur Loris <sup>a</sup> Potto <sup>a</sup> Tarsier <sup>a</sup> Galago <sup>b</sup>			Orangutan <sup>a</sup>	
Mated pair with offspring	Hapolemur Indri Avahi Galago <sup>b</sup>	Aotus Callicebus Marmosets <sup>a</sup>		Gibbon	"Family"
One-male group			Patas Theropithecus gelada <sup>a</sup> Papio hamadryas <sup>a</sup>		
Troop, but oriented to one male		Spider monkey <sup>a</sup>	Cercopithecus ascanius Colobus <sup>a</sup> Langurs <sup>a</sup>	Gorilla	
Troop (multiple adults of both sexes)	Lemur macaco Lemur catta Lemur fulvus Propithecus verreauxi	Howler Squirrel monkey <sup>a</sup> Capuchin <sup>a</sup> Spider monkey <sup>a</sup>	Macaques Savannah baboon Vervet Mangabeys <sup>a</sup> Colobus Langurs <sup>a</sup>	Chimpanzee <sup>a</sup>	"Band"
Unstructured aggregation	Galago <sup>b</sup>		Theropithecus gelada <sup>a</sup> Papio hamadryas <sup>a</sup>	Chimpanzee <sup>a</sup>	"Dialect group"

a. Inadequate field data.

b. Changes during annual cycle.

c. Changes during daily cycle.

d. Geographical variation.



mate species young adult males may change groups or live for a time in an all-male group or as solitary animals. But most remain in their natal group, where mutual tolerance and group cohesion are accomplished by such different behavior patterns as the dominance hierarchy, the persistence of mother-offspring bonds into adult life, an elaborate repertoire of affective communication, and the conservative tradition of the group. In such groups each individual has a status with regard to every other individual, and through the long period of infancy and adolescence, the individual learns to anticipate the responses of other group members and behave accordingly. It is in the rich behavioral context of these Old World primate social groups that many of the behaviors which we have come to consider "human" were first developed.

In patas monkeys (*Erythrocebus patas*), hamadryas baboons (*Papio hamadryas*), and geladas (*Theropithecus gelada*) the basic social unit comprises several adult females and their offspring but only one adult male. Still another primate grouping pattern is the mated pair with offspring, reported to be found in the Indridae, in some small South American monkeys, and in the gibbon. This form of social organization has been compared to the human family, but it is actually the result of very different behavior patterns. An adult gibbon, for example, is antagonistic toward all other gibbons except its mate and its immature offspring. As a result, each gibbon pair lives apart from all other pairs, and even their own offspring are apparently driven away at maturity. In this respect the gibbon group is more comparable to a mated pair of birds during the breeding season than it is to the human family. (Unlike most birds, the gibbon group is stable and unaffected by seasonal variation.) The human family cannot occur as an isolated pair of adults but is always a subgroup within a larger social unit.

Most of the prosimians are small, arboreal, nocturnal creatures which, like many other such mammals, forage as individuals, coming together only rarely. Yet some, such as *Propithecus verreauxi* and some species of lemurs (such as *Lemur macaco*), live in small troops. The small galago or bushbaby (*Galago senegalensis*) forages singly at night but may sleep in small groups during the day and form mated pairs during part of the year. (The most complete discussion of the classification, social groups, and activity patterns of Madagascar lemurs is in Petter 1962 and is summarized in Buettner-Janusch et al. 1962 and in DeVore 1965a.)

Still another form of social group is described for the chimpanzees by Goodall (1965) and by

Reynolds and Reynolds (1965). Small subgroups of a population, such as females with young, adult males, and mixed groups of males and females, roam freely in a local area of forest. These clusters are in frequent contact and may gather together, for example, at food trees but do not appear to be organized into large persistent social units. On the other hand, recent observations of chimpanzees by the Japan Monkey Center personnel suggest a group size and organization much like a monkey troop: when crossing open country between forest patches, the chimpanzee group consistently numbered 43 and contained a nucleus of adult males, females in estrus, and females with infants. On the basis of brief observations some monkey species, especially in the New World, may also live in loose aggregations. The aggregations of hamadryas baboons are different in that the one-male units remain intact even though more than seven hundred individuals may gather at the same sleeping place (Kummer & Kurt 1963).

As a result of field studies in progress, it will soon be possible to refine the descriptive categories shown in Table 1 and to add many new species. While the table reveals certain trends, such as the solitary life of many nocturnal prosimians and the frequent occurrence of the "troop organization" in monkeys, it also illustrates the difficulty of abstracting a model of "monkey" or "ape" social organization. Further, gross categories such as those in Table 1 should not obscure major behavioral differences: the frequency, intensity, and complexity of social contacts among members of a lemur troop are far fewer than comparable contacts among members of a monkey troop. This list of primate species could be tabulated with respect to characteristics such as habitat, diet, and group size (Crook & Gartlan 1966). For example, with the possible exception of some prosimians, males of all these species are dominant over females, and this is broadly reflected in the degree of sexual dimorphism they display. There is great variation in this characteristic, ranging from the gibbon, in which the sexes are scarcely distinguishable in size, to the baboon and the gorilla, where the male may be two or three times as large as the female. Sexual dimorphism with respect to body size, size of canine teeth, and general robustness of the skeleton and musculature all tend to correlate with the extent to which a species is adapted to the ground. Indeed, the degree to which an Old World primate species is adapted to life in the trees or to life on the ground is one of the best ways of predicting the size of the group, the degree of sexual dimorphism, the intensity of dominance relations among

adults, the use of loud vocalizations in intergroup spacing, and the size of the area over which the social group ranges ( DeVore 1963 ).

### Dominance and aggression

"Dominance" has been defined in many ways by different workers, and further study will undoubtedly show that quite different behaviors have been subsumed under this rubric. Common to all concepts of dominance, however, is the notion that the dominant individual can assert priority over a subordinate in order to gain some desired objective. Although expressions of dominance vary widely from one species to another, between groups within the same species, and even in the same individual at different times, it is clear that dominance interactions are fundamental to primate behavior and that they are especially prominent in the Old World monkeys, apes, and man. It is in the primarily ground-dwelling species such as baboons and macaques that dominance behavior is most apparent, and it is in the social groups of these species that dominance hierarchies exert the most influence in social organization. In these monkeys an adult male is about twice the size of an adult female, and there is typically a pronounced gap between the bottom of the male dominance hierarchy and the top of the female hierarchy. In more arboreal species, females are usually more nearly the size of males, and the dominance hierarchies of the sexes may overlap to some extent. In such species expression of dominance is often less frequent and more subtle, and mere avoidance, for example, may be as important as a challenge or attack in the demonstration of dominance status. In one-male groups the male is larger and more dominant than the females and does not tolerate any other adult males. In the loose aggregation of a species like the chimpanzee, there are frequent dominance interactions between adults, but it has not been reported whether these result in a hierarchy such as gorilla groups show.

The rank of any individual in a dominance hierarchy was formerly considered to be the result of the success of that individual in dominance interactions with each of the other individuals in the group. However, from recent studies of Old World monkey and ape groups, it appears that dominance status is seldom, if ever, simply the outcome of an interaction between two individuals. Any of a series of circumstances may intervene in the free expression of dominance between two individuals. The mother comes to the aid of her offspring during childhood and, at least among rhesus monkeys, into adult life as well ( Sade 1965 ). Other individ-

uals frequently join the combatants when a fight starts, and adults who tolerate each other will often support each other when one of them is challenged or attacked ( DeVore 1965b ). Thus, an individual's dominance status is the result of a variety of factors, including past success in dominance encounters, the status of its mother, and the number and status of the individuals who will support it when challenged.

**Threat behavior.** Baboons and macaques, among whom thresholds of aggression are lowest and agonistic encounters most frequent, exhibit the most elaborate repertoire of gestures and vocalizations conveying anger, threat, and fear. One form of threat behavior which is common in all the Old World primates ( including man ) is the "redirection of aggression," in which one animal threatens another and the threatened animal redirects the threat to a third individual or even to an inanimate object such as a stone or branch. For example, the loser of a fight between two adult males may chase a more subordinate male or a female or he may shake a tree limb vigorously. If he has chased a female, she will often chase or threaten a more subordinate female or a juvenile. This chain reaction of aggressive acts probably serves to periodically reinforce the various dominance statuses of all the participating group members.

**Male dominance.** Dominance relationships among adult males in baboon and macaque groups are often stable over many months or even years. A male's status is ultimately based on his strength and fighting ability, and unless the group is very large, the males tend to be arranged in a linear hierarchy, or "peck order," in which each male is dominant over all individuals beneath him. This stable hierarchy of recognized status serves to reduce the number of agonistic interactions among males. In some savannah baboon, rhesus monkey, and Japanese macaque troops, several males in their late prime may be so tolerant of each other that dominance between them is rarely expressed, and in such troops these males may constitute a nucleus in which the members support each other in dominance interactions. As a result, these older males, some of whom would lose a fight to a younger male in individual combat, are able to hold the young males in check by combined action ( DeVore 1965a, pp. 54-71; Imanishi 1960 ). It is this nucleus of older supporting males that takes the initiative in protecting the group from outside danger, in repelling individuals who try to join the group, and in breaking up fights among group members. On the contrary, among gorillas and in small monkey groups this initiative is taken by a



single dominant, "alpha" male. Dominance behavior is difficult to understand in captivity, where it seems to be disruptive and antisocial. But in natural surroundings, dominant males protect weaker group members, and females and juveniles seek out the dominant males to sit beside or to groom. So, although the dominance hierarchy depends ultimately on force, it leads to relative peace and stability within the group.

**Female dominance.** The dominance hierarchy of females tends to be unstable and to depend to some extent on the female reproductive cycle. Females in estrus are more active and aggressive and are frequently under the protection of a consort male. Among baboons and macaques, adult males are extremely protective of young infants, and this protection also extends to the mother as long as she is with a young infant. Thus, although it is clear that some adult females are more dominant than others in almost all circumstances, the interference of reproductive cycles in female-female dominance interactions seems to preclude the establishment of stable, linear dominance hierarchies.

### Sexual behavior

As in humans, most female monkeys and apes have a menstrual cycle of about thirty days. Because humans do, and monkeys in captivity may, breed at all times of the year, the belief has been widely held that Old World primates show no seasonal variation in mating activity. Since the publication of Zuckerman's *The Social Life of Monkeys and Apes* (1932), it has often been asserted that the basis of monkey and ape societies is sexual attraction; however, we now know that the nonhuman primate group performs many functions besides providing its members with sexual outlets.

There are many differences in patterns of sexual behavior between man and the nonhuman primates. In the first place, in at least some species of macaques, there are sharply defined mating and birth seasons. The data are best for the Japanese macaque and for the rhesus on Cayo Santiago, where records have been kept longest. Lancaster and Lee (1965) show that Old World monkeys tend to have at least a birth peak, if not a sharply defined birth season. Most monkeys give birth every year, and as in many birds and mammals, their reproductive cycles are presumably affected by such phenomena as day length and temperature. The reproductive cycle of some species, such as the baboons and the apes, is longer than a year, and if marked seasonality of sexual reproductive behavior

exists in these species, it has not yet been demonstrated.

Another reason that opportunity for mating is an inadequate explanation of social cohesion in nonhuman primate groups is the fact that the female is sexually receptive for a relatively small portion of her life. Females will accept a male only on those days of their sexual cycle when they are in estrus, and sexual cycles are inhibited during pregnancy and lactation. Females and their young constitute the core of monkey groups, and yet the amount of sexual activity in the life of any female, or within any group during the year, is very small. In a species with a mating season, disruptions are most common during this season. At this time extragroup males attempt more actively to join the group, and the existing competition among males already in the group may become more intense. Koford (1966) reports that on Cayo Santiago 66 per cent of all rhesus males who leave a group do so during the mating season. One of the Cayo Santiago groups subdivided into five groups between 1958 and 1964, each subdivision occurring during the mating season. On the other hand, social bonds within groups are probably strongest during the birth season, when mothers are clustered together with their new infants and adult males are actively caring for the infants of the previous year.

**Mating patterns.** In such species as the gibbon and the patas monkey (and in small groups of any species), only one adult male does the mating, but in most Old World primate groups the female is potentially accessible to every male in the group during some part of her estrous cycle. In such dominance-oriented species as baboons and macaques, the tendency is for the most dominant male to have priority of access to an estrous female, but his mating activity is usually confined to her period of maximum receptivity. For example, a female baboon is receptive for about 12 days in a thirty-day sexual cycle; the degree of receptivity can be gauged by the amount of tumescence in the sexual skin of the perineal area. During the early period of swelling she is mounted by juvenile and subordinate adult males, and it is only when tumescence is at a maximum that she is sought out by the dominant males. If there are a number of adult males of approximately equal dominance and only one receptive female, the males may threaten and fight for her possession. Although copulations early in her cycle are casual and transitory, dominant adult males usually form a consort pair with a receptive adult female. The pair may remain together for several days, or if there is competition among the

males, the pairing may be interrupted after only a few hours. A baboon male's ability to maintain possession of an estrous female when he is being threatened will depend in part on the support he can expect from other males (DeVore 1965b). The female must cooperate with the male during copulation, and she is thereby able to exercise some personal preference and to take some initiative in making sexual contacts. Thus, mating patterns in even a sexually dimorphic species like the baboon are more than a simple reflection of the male dominance hierarchy. There are opportunities for subordinate males to copulate before the female becomes attractive to the dominant males; a male's ability to mate may depend on his relationship to other, supporting males; and the female herself is able to exert some initiative.

### Prolonged immaturity

By comparison to other mammals of similar size, the period of infant dependency is remarkably long in primates, and the necessity for providing care and protection of immature offspring is a dominant feature of monkey and ape social life. Ordinarily an infant in danger is retrieved by its mother, but adult males of many varieties of monkeys will threaten, distract, or attack a potential predator. Aggression toward predators is most prominent in ground-dwelling species such as baboons and macaques; for example, on an average day a baboon group may range as far as two or three miles from the safe refuge of trees.

The birth of an infant not only changes the quantity and quality of its mother's social interactions, but every member of the group is affected in some measure. Adults and juveniles are highly attracted to young infants and repeatedly approach them to inspect, fondle, groom, and play with them. The newborn infant primate clings to its mother at all times, imposing substantial behavioral constraints not only on her but on the entire group. This prolongation of preadult life is so biologically expensive for the species that there must be major compensatory advantages. In a recent discussion, Washburn and Hamburg (1965) have suggested that this prolonged period of immaturity makes it possible for the species to adapt to a wide variety of environmental contexts because of the opportunity it provides for complex learning in the young. A large number of laboratory and field studies support the fact that much of the normal adult behavior of the species must be learned. In nature immature monkeys are in frequent contact with adults of both sexes and with their peers. The

major activity of the young primate is "play," and it is during these long hours of daily play that the juvenile learns both the features of its physical environment and the appropriate responses to the other animals in its social environment. The development of normal sexual behavior, the expression of effective patterns of affection and dominance, and the ability to rear an infant competently are all social skills which must be learned, and monkeys raised in isolation show major deficits in all of these behaviors (Harlow & Harlow 1965; Mason 1965). Social deprivation experiments also indicate how even less drastic events in the early life of a young primate reared under normal conditions may lead to the striking individual differences in adult temperament and social skills that are so frequently reported by research workers.

**Mother-offspring relations.** It has long been obvious to observers that the relationship between a mother and her offspring is the most intense and prolonged in primate life. During the usual field study it is impossible for the observer to know what genealogical relationships may exist between adults or between any group members who are beyond the weanling stage. Rejection by the mother when she weans her infant seems so firm, and the effect on the infant so traumatic, that the tendency has been to assume that their early close bonds have been severed permanently. (Moreover, the mother's attentions are soon redirected to the care of another newborn infant.) But these conclusions are clearly incorrect, at least for some species, as prolonged observation of identified individuals has shown. Studies reported by Imanishi (1960) and Koford (1963) indicate that the rank of a young male in the dominance hierarchy is frequently very closely related to the rank of his mother; high-ranking mothers tend to rear high-ranking male offspring. The evidence is less clear for females, but it seems likely that high-ranking females may also be related as siblings or as mother and daughter.

Sade (1965) has offered the most detailed accounts of the importance of social bonds between uterine kin in rhesus monkeys. The most frequent subgroups are clusters of mature and immature offspring around an old female. These animals not only sit together and groom each other more often than they do other group members, but they frequently support each other in dominance interactions as well. Data indicating the importance of genealogically related subgroups are most convincing for the Japanese macaque and the rhesus on Cayo Santiago, where individual records have been



kept for many years, but Jane Goodall (1965) has also described similar clusters of offspring around old female chimpanzees. Troops of Japanese macaques and rhesus monkeys which contain no adult males (except during the mating season) have been reported, but they are rare.

The achievement of high rank by the offspring of dominant females appears to be caused both by childhood experience and by the mother's support of her mature offspring. Infants of dominant females stay with their mothers and other dominant individuals near the center of the group, and in the Japanese macaque they may never go through the period of social peripheralization that is characteristic of most young males. They are born in the middle of the group, grow up there, and remain there as adults, assuming at a young age the roles peculiar to the dominant males at the heart of the group. Also, in day-to-day agonistic interactions, a young male who is supported in aggressive encounters by a dominant mother will simply win more of these encounters than one whose mother is subordinate. Presumably the long period of conditioning during infancy and adolescence strongly reinforces a male's expectation of success or failure in the social dominance interactions of adult life. The question of the effect of genetic inheritance on the achievement of social dominance should be investigated; this could easily be done by cross-fostering experiments, in which the offspring of a subordinate female and a dominant female would be exchanged at birth.

Sade reports that rhesus mothers on Cayo Santiago continue to support their adult sons in dominance interactions with other group members. Furthermore, a mother ordinarily remains dominant over her son all of her life. Apparently correlated with this behavior is the fact that copulation between mother and son is very rare. In one of the few known cases of mother-son incest, the two had engaged in a vicious fight, and the son had subsequently become dominant over his mother before the mating occurred. Although they are not yet well understood, it appears that the complicated patterns of nurturance and dominance behavior between mother and son may inhibit copulation between them and that the aversion to incest may be present among members of a number of Old World primate species.

**Adult male-offspring relations.** Paternity in nonhuman primate groups is rarely known or knowable either to the investigator or to the monkeys. Adult males seek to protect *any* infant from injury either from group members or from predators. However, in some species an adult male may

form a special protective relationship with a particular infant. In the Japanese macaque, for example, males frequently establish such a relationship with yearlings when these are rejected by the mother and replaced by a new infant (Itani 1959). Analogous behavior has been reported in other macaques and baboons (Washburn et al. 1965). Among some New World marmosets a male may carry an infant at all times except when the mother is nursing it. The one-male unit of the hamadryas baboon develops because young adult males forcefully adopt immature females, "mothering" the females until they reach sexual maturity (Kummer 1968).

There are no indications from field studies that the adult male in *arboreal* Old World primates establishes strong social bonds with infants or juveniles beyond protection of infants when the group is threatened by a predator. Even in the ground-adapted patas monkey and in the vervet monkey, the role of the male seems limited to protection or to distracting predators, and the present evidence indicates that more elaborate male-offspring relations arise only in terrestrial primates living in large social groups.

### Human and primate social organization

Only a few years ago nonhuman primate behavior was assumed to be simple, stereotyped, and easy to observe; with notable exceptions (Carpenter 1964), field reports were largely anecdotal accounts by casual observers. Observations of zoo colonies suggested that primate groups were a disorganized rabble, tyrannized by aggressive males and held together only by sexual attraction. From the perspective of modern field studies, the conclusions based only on the behavior of primates in captivity bore as little relation to the social organization of free-ranging groups as would a monograph on middle-class society based solely on a study of inmates in a maximum-security prison.

Social organization is not identical for any two primate species. Human social organization, with its varied cultural manifestations, is unique; but studies of monkeys and apes are helping to distinguish those behavior patterns that are truly unique from those which man shares with his nearest relatives. Before the urban revolution human social organization was centered on small, face-to-face groups of adults involved in the quest for food and the protection of immature offspring, and it is not surprising that patterns of dominance and aggression and of nurturance and affection seem so similar in men, monkeys, and apes. The division of labor by sex, with females specializing in child

care and males as protectors, and the performance of different social roles according to status are familiar patterns to the sociologist. The evolutionary basis of man's unique biological and social system is treated elsewhere in this encyclopedia, but it seems appropriate to outline here some of the features of human organization which seem to distinguish men most clearly from the other primates [see EVOLUTION, article on HUMAN EVOLUTION].

Much of the contrast between human and primate social organization is based on human language. Communication in monkeys and apes is a rich complex of vocal, gestural, and tactile signals, but the majority of these signals are reflections of changing emotional states among the communicants and are comparable to the communication systems of other animals, not to language [see COMMUNICATION, ANIMAL; see also Marler 1965; Bastian 1965; Altmann 1967]. Language is so fundamental to human life that it is impossible to dissociate it from other distinctly human behaviors, such as toolmaking and the willingness to share food. Nonhuman primates do not share food, even with their own infants, but neither do they have tools to gather food or an improved site or camp from which they can disperse during the day in complementary gathering and hunting activities and where they can convene again to share the results of the day's quest.

The most distinctive feature of human social organization, the family, is adaptive only in the context of an economy based on tool use and sharing; presumably it also requires other uniquely human traits such as the ability of the female to mate at all times. On the other hand, many behavior patterns that have been viewed as ramifications of the human family are actually present in the Old World primates—the maternal affectional system and its extension to uterine kin, the close association of siblings, the protection and nurturance by adult males of the young—and it is no longer so difficult to understand how the human social systems could have evolved. With language, an incest aversion between mother and son became an incest taboo; with language, the already close ties among uterine kin were explicitly recognized and socially extended.

Field and laboratory studies of primates were not begun on a large scale until the 1960s, but they are now attracting the attention of persons in biology, ethology, and all branches of the behavioral sciences. Techniques of observation and analysis are constantly being refined, and behavioral studies are being combined with experiments in pharma-

cology, hormonal studies, and neurophysiology (Washburn & Hamburg 1968). For example, miniature telemetering devices implanted in the animal have been used to record and direct the behavior of individuals in a free-ranging group. Ethological principles and techniques of observation are being applied to humans and the results compared to studies of nonhuman primates [see ETHOLOGY]. There is every indication that in the next few years this renewed interest in the nonhuman primates will yield results of increasing relevance to the social sciences.

IRVEN DEVORE

#### BIBLIOGRAPHY

- ALTMANN, STUART A. (editor) 1967 *Social Communication Among Primates*. Univ. of Chicago Press.
- BASTIAN, JARVIS R. 1965 Primate Signalling Systems and Human Languages. Pages 585-606 in Irvén DeVore (editor), *Primate Behavior: Field Studies of Monkeys and Apes*. New York: Holt.
- BUETTNER-JANUSCH, JOHN (editor) 1963-1964 *Evolutionary and Genetic Biology of Primates*. 2 vols. New York: Academic Press.
- BUETTNER-JANUSCH, JOHN et al. (editors) 1962 The Relatives of Man: Modern Studies of the Relation of the Evolution of Nonhuman Primates to Human Evolution. New York Academy of Sciences, *Annals* 102: 181-514.
- CARPENTER, CLARENCE R. 1964 *Naturalistic Behavior of Nonhuman Primates*. University Park: Pennsylvania State Univ. Press.
- CLARK, W. E. LeGROS 1960 *The Antecedents of Man: An Introduction to the Evolution of the Primates*. Chicago: Quadrangle.
- CONFERENCE ON SEX AND BEHAVIOR, BERKELEY, CALIFORNIA 1965 *Sex and Behavior*. Edited by Frank A. Beach. New York: Wiley.
- CROOK, J. H.; and GARTLAN, J. S. 1966 Evolution of Primate Societies. *Nature* 210:1200-1203.
- DEVORE, IRVEN 1963 A Comparison of the Ecology and Behavior of Monkeys and Apes. Pages 301-319 in S. L. Washburn (editor), *Classification and Human Evolution*. Chicago: Aldine.
- DEVORE, IRVEN (editor) 1965a *Primate Behavior: Field Studies of Monkeys and Apes*. New York: Holt.
- DEVORE, IRVEN 1965b Male Dominance and Mating Behavior in Baboons. Pages 266-289 in Conference on Sex and Behavior, Berkeley, Calif., *Sex and Behavior*. Edited by Frank A. Beach. New York: Wiley.
- DEVORE, IRVEN; and LEE, RICHARD 1963 Recent and Current Field Studies of Primates. *Folia primatologica* 1, no. 1:66-72.
- EIMERL, SAREL; and DEVORE, IRVEN 1965 *The Primates*. New York: Time Inc.
- GOODALL, JANE 1965 Chimpanzees of the Gombe Stream Reserve. Pages 425-473 in Irvén DeVore (editor), *Primate Behavior: Field Studies of Monkeys and Apes*. New York: Holt.
- HARLOW, HARRY F.; and HARLOW, MARGARET K. 1965 The Affectional Systems. Volume 2, pages 287-334 in Allan M. Schrier, Harry F. Harlow, and Fred Stollnitz (editors), *Behavior of Nonhuman Primates: Modern Research Trends*. New York: Academic Press.



- IMANISHI, KINJI 1960 Social Organization of Subhuman Primates in Their Natural Habitat. *Current Anthropology* 1:393-407.
- ITANI, J. 1959 Paternal Care in the Wild Japanese Monkey, *Macaca fuscata fuscata*. *Primates* 2, no. 1: 61-93.
- JAY, PHYLLIS 1965 The Common Langur of North India. Pages 197-249 in Irven DeVore (editor), *Primate Behavior: Field Studies of Monkeys and Apes*. New York: Holt.
- JAY, PHYLLIS (editor) 1968 *Primates: Studies in Adaptation and Variability*. New York: Holt.
- JOLLY, ALISON 1967 *Lemur Behavior: A Madagascar Field Study*. Univ. of Chicago Press.
- KOPFORD, CARL B. 1963 Rank of Mothers and Sons in Bands of Rhesus Monkeys. *Science* New Series 141: 356-357.
- KOPFORD, CARL B. 1966 Population Changes in Rhesus Monkeys: Cayo Santiago 1960-1964. *Tulane Studies in Zoology* 13, no. 1:1-7.
- KUMMER, HANS 1968 Two Basic Variations in the Social Organization of the Genus *Papio*. In Phyllis Jay (editor), *Primates*. New York: Holt.
- KUMMER, HANS; and KURT, F. 1963 Social Units of a Free-living Population of *Hamadryas Baboons*. *Folia primatologica* 1:4-19.
- LANCASTER, JANE B.; and LEE, RICHARD B. 1965 The Annual Reproduction Cycle in Monkeys and Apes. Pages 486-513 in Irven DeVore (editor), *Primate Behavior: Field Studies of Monkeys and Apes*. New York: Holt.
- MARLER, PETER 1965 Communication in Monkeys and Apes. Pages 544-584 in Irven DeVore (editor), *Primate Behavior: Field Studies of Monkeys and Apes*. New York: Holt.
- MASON, WILLIAM A. 1965 The Social Development of Monkeys and Apes. Pages 514-543 in Irven DeVore (editor), *Primate Behavior: Field Studies of Monkeys and Apes*. New York: Holt.
- PETTER, J. J. 1962 Recherches sur l'écologie et l'éthologie des lémurs malgaches. Musée National d'Histoire Naturelle, *Mémoires Série A* 27, no. 1:1-146.
- REYNOLDS, VERNON; and REYNOLDS, FRANCES 1965 Chimpanzees of the Budongo Forest. Pages 368-424 in Irven DeVore (editor), *Primate Behavior: Field Studies of Monkeys and Apes*. New York: Holt.
- SADE, DONALD S. 1965 Some Aspects of Parent-Offspring and Sibling Relations in a Group of Rhesus Monkeys, With a Discussion of Grooming. *American Journal of Physical Anthropology* New Series 23:1-17.
- SCHALLER, GEORGE B. 1963 *The Mountain Gorilla: Ecology and Behavior*. Univ. of Chicago Press.
- SCHRIER, ALLAN M.; HARLOW, HARRY F.; and STOLLNITZ, FRED (editors) 1965 *Behavior of Nonhuman Primates: Modern Research Trends*. 2 vols. New York: Academic Press.
- SOUTHWICK, CHARLES H. (editor) 1963 *Primate Social Behavior; an Enduring Problem: Selected Readings*. Princeton, N.J.: Van Nostrand.
- STRUBSACKER, T. 1965 Behavior of the Vervet Monkey, *Cercopithecus aethiops*. Ph.D. dissertation, Univ. of California.
- SYMPOSIUM ON THE PRIMATES, LONDON, 1962 1963 *The Primates*. Edited by John Napier and N. A. Barnicot. The Proceedings of the Symposium of the Zoological Society of London, No. 10. London: The Society.
- WASHBURN, S. L. (editor) 1963 *Classification and Human Evolution*. Chicago: Aldine.
- WASHBURN, S. L.; and HAMBURG, DAVID A. 1965 The Study of Primate Behavior. Pages 1-13 in Irven DeVore (editor), *Primate Behavior: Field Studies of Monkeys and Apes*. New York: Holt.
- WASHBURN, S. L.; and HAMBURG, DAVID A. 1968 Aggressive Behavior in Old World Monkeys and Apes. In Phyllis Jay (editor), *Primates*. New York: Holt.
- WASHBURN, S. L.; JAY, PHYLLIS C.; and LANCASTER, JANE B. 1965 Field Studies of Old World Monkeys and Apes. *Science* New Series 150:1541-1547.
- ZUCKERMAN, SOLLY 1932 *The Social Life of Monkeys and Apes*. New York: Harcourt.

## III

## THE REGULATION OF ANIMAL POPULATIONS

Ecologists have been studying the dynamics of animal populations for roughly forty years, trying to discover how their numbers are regulated. The subject is of much practical importance in a world in which man, either knowingly or unknowingly, interferes drastically in this process and in which our own exploding population has become a central threat to human survival. Scientific knowledge in this field has reached a stage at which generalizations are possible, although final agreement has not been reached.

Populations of animals obviously cannot all have the same life-cycle characteristics. Many of the insects run through several generations in a single summer, after which the population remains dormant during the winter and regenerates in the spring. Larger forms of animals, on the other hand, tend to live longer, and most of them have populations that keep going all the year round. Some populations fluctuate in numbers from season to season; others remain practically constant from year to year, even from decade to decade. Striking examples of the fluctuating kind include the arctic lemmings, many of the northern game birds and fur bearers, and the subtropical locusts. The steady state typically applies to most kinds of big game, many birds, and other animals both large and small, especially in temperate and tropical regions with equable climates. There is a continuous gradation between the extremes, and under natural conditions the populations of most animals in most situations resemble one another in one very important respect, namely, that there is no long-term trend either up or down.

The normal situation is one in which the density of population moves or fluctuates irregularly, either weakly or strongly depending on circumstances, to a varying distance on either side of a constant average value. If and when permanent changes

take place in the environment, the average is also likely to shift and will assume a new level once the environmental change is complete. Therefore, we need to find out what determines the average, what causes the fluctuations and prevents them from swinging out of control in one direction or the other until the population either approaches infinite size or falls to zero.

**Population income and loss.** Animals must in general possess the fecundity necessary to augment their numbers; otherwise they would quickly die out. In the course of time local populations receive an *income* of new recruits, largely as a result of reproduction, although some may come through immigration. They are also subject to *loss*, some of it by emigration and the rest through mortality. If the population stays roughly constant in number over the years, it means that the cumulative totals of recruitment and loss have balanced out.

The fact that populations tend to maintain a constant average level and that when fluctuations occur they do not proceed very far in one direction or the other without being checked indicates that there must be some stabilizing process at work. For example, if the recruitment rate were always 20 per cent per annum and the death rate 25 per cent, the population would steadily diminish to the vanishing point; similarly, if the two rates were reversed, it would expand at 5 per cent compound interest. But let these rates vary according to whether the population density is high or low—so that high density produces low recruitment or high mortality or both, and low density the opposite—and then any continuous trend of numbers up or down can always be prevented. The phenomenon typically found in nature, of population density balanced within relatively narrow limits of variation, cannot in fact be accounted for in any other way than by regulators that operate in this manner and are “density dependent” in their action.

There is one important school of thought which believes that density-dependent action arises mainly from the factors causing losses in the population. According to this hypothesis, animals are supposed always to reproduce as fast as they efficiently can, the surplus so produced being removed in a density-dependent manner by the various destructive agencies, namely predators, parasites and disease, starvation, and accidents of the physical world, such as floods and droughts. On the face of it, there seems nothing impossible or improbable in this: disease, for example, would be expected to spread fastest in overcrowded conditions, and starvation would have a similar incidence. The

rate of accidental mortality seems perhaps the factor least likely to have any connection with population density.

There are, however, many large animals, of which primitive man was a good example and lions and eagles are others, which effectively have no predators and appear not to be regularly subject to death from disease on a significant scale, except perhaps among old or outcast members of the population. We find ourselves, therefore, forced back on starvation as the only remaining density-dependent agency that might be capable of removing a population surplus. This possibility calls for close examination.

Population density clearly has a very special relationship to food supply. Animals “select” their environments or habitats in accordance with their natural inclinations. The habitat selected has to provide them with the resources they need for sustaining life, including food of adequate quantity and quality, space, shelter, and breeding sites. In the great majority of species, food is the resource that sets the upper limit on population density, and determines the “carrying capacity” of the habitat. Habitats can therefore generally be graded as richer or poorer almost entirely in terms of their food productivity, and the average population density of animals living in them has been shown, in situations where both can be measured, to be closely related to it.

However, the correlation between food supply and population density has no direct connection with starvation. The normal condition is that all individuals living in a particular habitat get enough to eat. If starvation occasionally occurs, it almost always comes as an emergency resulting from an accidental cause, such as exceptional cold or drought, and is not in any way dependent on the density of population that happens to be affected by it.

**Food resources harmed if overtaxed.** Human experience in exploiting purely natural resources of animal and plant life, whether they happen to be the virgin pastures of the prairies or fish or game, shows that most of them can all too easily be overtaxed. Persistently overgrazed pasture quickly deteriorates and especially in arid climates can be reduced to a worthless desert in a few centuries. If we take too big a crop from our fisheries, whether for commerce or sport, or press too hard on fur bearers, wild fowl, or game, the stocks and the yields decrease. Unfortunately, long after the demand has become excessive, it is almost always still possible to go on making money, and we tend therefore to shut our eyes to the in-



creasingly disastrous consequences piling up for the future. This highly characteristic danger inherent in the exploitation of "recurrent resources" has often led in the past to the decimation of the more lucrative and vulnerable prizes of the hunter, such as the passenger pigeon, the northern right whale and the blue whale, the fur seal and marine turtle, or the rhinoceros of Africa and Asia, now threatened because their horns are highly valued by people in the Far East for their supposed aphrodisiac powers.

In exploiting their own sources of natural food, animals face exactly the same dangers as man. For the same reason, therefore, they must exercise restraint while there is plenty, if that plenty is to be conserved and future yields sustained.

This is a deduction of great importance, for it implies that starvation can be ruled out as a practicable means of curbing population growth. Long before starvation occurred, disastrous damage would have been done to the food resource by over-exploitation, and the survival of the entire population would be threatened.

It can be concluded, therefore, that not only are there many kinds of animals which are not subject to predation or disease on an effective scale, but that starvation can also be eliminated as a practical regulating factor, because damage to the habitat is a precondition and, consequently, constitutes a continual threat to the safety of the group as a whole.

### Regulation through "conventions"

It is now necessary to explore the alternative hypothesis that animals themselves possess the means of controlling numbers and population density. If they were able to do this, it would obviously be an adaptation of great value, because an ability to control population density would imply that it could be adjusted to some particularly favorable level, and that level could very well be the optimum one, as far as exploiting the food resource is concerned. Ideally, it would then be possible to take the maximum sustainable crop from the food resource and always feed the largest number of mouths that could be safely allowed without ever running the risk of overtaxing or depleting the supply.

It is a common situation for an animal population to depend for a period of weeks or months on a standing crop that is not at the same time being replenished. This applies, for example, to birds feeding on seeds and berries in the fall or hunting for hibernating insects during the winter. When they first broach the food stock, it is at

maximum abundance and for perhaps hours or days could provide food for a prodigious number of exploiters. But it must be made to last much longer than that, and somehow the number or population density of exploiters must be restricted from the outset. The food itself will not restrict them, because it is superabundant; and yet the reason for imposing the restriction is related to food and food alone. What is required is some artificial method of limiting the population density, one that is, if possible, closely correlated with food production, serving as a buffer or shield in preventing overexploitation: something, in fact, closely analogous to the restrictive conventions entered into by states and nations to prevent commercial overfishing in the sea.

It is at once apparent that a great many conventions of this kind exist in the natural world. The best known of them is the territorial system adopted by many kinds of birds during the breeding season. In the case of robins, cardinals, song sparrows, and a host of other birds, the males of a species compete with each other, each claiming and defending a territory on which pairing, nesting, and rearing of a young family subsequently take place. The males will not tolerate the crowding of territories beyond a certain limit, so that once a particular habitat has been taken up in claims, no further males can get in; any that attempt it are soon forced to leave and try somewhere else.

This is a perfect example of a conventional density-limiting mechanism, and one easily understood. What has happened is that direct contest between rival members of the group for actual food has been sidetracked into equally furious competition for a piece of ground—a conventional substitute. Once the ground is "owned," the owner and his dependents can use the food that the territory contains in perfect freedom, without further dispute, and provided the conventional territory size is large enough, there need be no danger that their demands will overtax the supply.

In a short article one cannot hope to do more than suggest the broad implications of this hypothesis. It must suffice to say that many of the density-limiting conventions that can be observed in nature are less direct or more sophisticated than the simple territorial one. For example, birds that depend on the resources of the ocean cannot stake individual territories for gathering or nesting on the sea itself. Instead, they breed in a colony on shore which in most cases is quite artificially limited in size and in which each successful pair gains a token territory in the form of a small nest

site. Adults that have not succeeded in winning a nest site take on the status of onlookers and are frequently found at seabird colonies. They are evidently inhibited from nesting outside the traditional boundary or from starting a new colony around the corner. This conventional recognition of a colony is presumably related to the food supply, limiting the local population to a size that will be permanently immune to overfishing; however, as compared to the territorial system, the form it takes is more abstract or artificial because the object competed for is reduced to a few square feet or a hole or crevice, rather than the entire food-supplying territory. Abstraction can go even further than this. In some cases it is only the right to membership in a group that is recognized; the alternative is rejection as an outcast. Thus, there may be no contest for actual property at all, but only for individual or personal status.

These conventions are normally concerned with competition between the members of the population of a single species of animal: that is to say, with intraspecific competition. This competition provides conventional goals or rewards that are substituted for actual food, and their effect is to limit the number (or density per unit of food) of members in the group and to unload any unwanted surplus to a safe distance.

It is important to note that the competition itself assumes artificial or conventional forms. Birds contesting for territories seldom fight and kill each other, although winning a territory can in fact easily be a matter of life or death as far as the respective consequences to the victor and vanquished are concerned; instead they sing and threaten by voice or display fine crests or splendid or startling patterns of plumage.

Social behavior provides the key. It takes but a short step at this point to reach the most striking and important generalization to emerge from the new hypothesis: Any group of individuals competing for conventional prizes by conventional methods automatically constitutes a society. Behind this there is the deeper implication that here is the essence or root from which all social behavior has stemmed: that the evolution of a primordial social organization occurred as an adaptive mechanism for the control of population density. A *society* could be defined as an organization capable of providing conventional competition among its members; in these terms, sociability is seen to have a biological basis. One does not need to ponder very deeply to see how closely this cap fits even in human social groupings, where leaders emerge even though individuals need not be engaged in

any direct form of competition. A society, whether human or animal, is in essence a brotherhood tempered with rivalry.

Inasmuch as it incorporates the mechanisms of competition, social organization appears to be indispensable in the process of maintaining an optimum population density. This is readily apparent in a territorial system, like that of birds, where neighbors of the same species are in constant social competition and by their behavior directly control the population density in the habitat. The long-term research on the Scottish red grouse being carried on near Aberdeen has shown, for example, that the grouse population on a heather moor consists of individuals all mutually known to their neighbors and greatly different in social standing. Dominant cock birds hold territories almost all the year round, covering the habitat like a mosaic; the most aggressive have the biggest territories. Socially subordinate cocks and unmated hens are often carried as members of the population although there is no room for them to establish any territorial claims for themselves. With the onset of winter, or when the productivity of the food supply declines for any other reason, it is the subordinates at the bottom of the social ladder that get squeezed out, leaving behind a surviving group adjusted in density to suit the lowered food level.

The social hierarchy, or peck-order system, thus works as a safety valve, or overflow mechanism, identifying and unloading those who cannot be supported when there is a drop in food supply. If the population were not thinned down in this way, damage would be done to the habitat by overtaxing the resource. By means of the social hierarchy not only is damage prevented but the remaining population is assured of an adequate diet and a reasonable chance of survival.

The hierarchy also operates in relation to breeding. Only established territory owners and their mates can breed, although unestablished non-breeding adults can still be carried in the population and form a valuable reserve to fill any gaps that appear in the ranks of the establishment. It may be noted that the hierarchy is another well-known biological phenomenon for which no general functional explanation has been given, although it finds an automatic place under the new theory.

Characteristically, red grouse compete under conventional rules, one of which is that, at least in the fall, the competition takes place only between dawn and about 8 A.M. After that the social groups flock amicably, and all feed side by side for the rest of the day. So fierce is the aggression among



the cock birds at dawn, however, that it is enough to subject some of the less-spirited competitors to a stress that not only drives them away from the habitat altogether but leaves them defenseless to predators and liable to succumb quickly to disease.

The convention of competing at dawn, and often at dusk, and leaving the rest of the day free for other necessary activities like feeding is exceedingly common and appears to shed a flood of light on the timing of such familiar phenomena as the cockcrow and the dawn chorus of birds, the dusk flighting of ducks, the massed maneuvers of starlings and blackbirds at their roosts, the synchronized choruses of some kinds of tropical bats and cicadas, and the activity of marine life such as the croaker and the snapping shrimp; there are innumerable other analogous phenomena in which an indication of present numbers seems to be provided through synchronized communal activity. Dawn and dusk are of course the two most easily perceived hours of the day, and this no doubt explains their wide use for brief social activities that call for a simultaneous general participation in order to achieve their effect.

**Homeostasis and epideictic feedback.** Community displays and contests provide the essential index of population density from day to day, producing "information" which can lead to an appropriate response when an adjustment to density is required. The whole process of balancing the population against the changing food productivity of the habitat appears to be a "homeostatic," or self-stabilizing, one, based on a feedback of information that sets the wheels of the adjusting machinery in motion whenever the balance has swung away from the optimal level. Community displays can not only lead to the expulsion of a population surplus, as in the case of the red grouse, but when the breeding season comes round, they can also serve to indicate how large a quota of recruits will be required to replenish the population or build it up to the particular level that is justified by the current carrying capacity of the habitat.

We need a technical term to describe this kind of activity, whose purpose is to sample population pressure, and "epideictic" has been chosen as having the right derivation and meaning. One could easily predict that epideictic displays would be especially conspicuous before the breeding season, when the population is about to be built up to its annual peak level; indeed, this is the time at which some of the displays already mentioned reach their height.

At that time in many species only the males are

required to take part in communal displays, each commonly acting as the representative of a mated pair; the females are otherwise occupied, frequently in more domestic duties. In these circumstances the males can be described as the "epideictic sex." The stimulation they receive from competing with rival males or sharing in exciting communal activities is density-dependent in its intensity: the more males there are, the hotter the pace. Rather frequently they assemble in unisexual swarms or groups, dancing in the air, perhaps, if they are flying insects or taking part in ritual tournaments, gymnastics, or parades, in the case of sage grouse and prairie chickens, tropical hummingbirds, manakins, and birds of paradise. They may have special vocal organs, as in songbirds, cicadas, frogs, drumfish, most crickets and katydids, howling monkeys, and many others. These are primarily used not to woo and win females, as was formerly thought, but to gain the prizes of social competition in the form of real estate and personal status. Exactly the same use applies to adornments, weapons, and scent glands. Much rethinking is therefore required on the hitherto vexed subject of sexual selection.

Epideictic displays are especially prominent among migratory animals as a prelude to their departure and as a means of smoothly handling heavy traffic over the vast tracts of country that often separate summer and winter quarters. This is especially true of birds. Without such displays the food resources would tend to be dangerously overtaxed in congested stopping places. The excited flocks of swallows and swifts, the spectacular massed maneuvers of locusts building up for a great emigration flight, and many of the big gatherings of birds, fruit bats, and insects to nightly roosts are associated with impending changes in population density through migration.

## Conclusions

In the light of this theory of self-regulation of animal populations effected through the agency of social conventions, we can turn back once more to the original questions and see how they have been answered. The average population level would appear to be set for most species by the size of sustained crop that the food resources can yield, and this level is actually achieved through a code of conventional behavior that prevents the growth of population from going above the optimum density. Some fluctuations in population level are produced by accidental events in the environment, including predators and disease, which may cause

extensive mortality; but in most cases they appear to arise from the operation of the homeostatic machinery, which automatically allows the population density to build up when resource yields are good and thin down when yields fall below average.

The homeostatic machinery responds directly to the pressure of social competition. The latter is part and parcel of the struggle for existence and depends in turn on food availability and the number of mouths to be fed. It mirrors (inversely) the standard of living prevailing at the time. In this, therefore, lies the answer to the question, What establishes the balance of numbers and prevents the progressive long-term increase or decline of populations? Fluctuations tend in practice to be checked by homeostatic processes, which can control the recruitment rate, the pressure to emigrate, and, sometimes on an immense scale, the amount of mortality resulting from stress. Only when there are long-term alterations in the environment itself will the average population levels undergo permanent change.

This sketch can give only an imperfect picture of the wide ramifications of the hypothesis. There is still one most important point to be mentioned, which is that man himself is conspicuously out of line with all the higher animals in being exposed to progressive population growth. The history of this seems perfectly clear. In many animals, including some of the mammals, fertility itself is under density-dependent control; overcrowding automatically leads to lowered, and finally suspended, recruitment. In primitive man, on the contrary, fertility was limited not through the internal effects of hormones on the rate of ovulation or on the resorption of embryos in the uterus (as in rabbits, foxes, deer, etc.) but through cultural conventions, customs, and taboos which imposed restrictions on sexual intercourse for mothers who already had a child at the breast or prescribed compulsory abortion, infanticide (especially of female infants), human sacrifices, or head-hunting expeditions. These had evolved in prehistoric man as part of a socially integrated control mechanism which kept populations nicely balanced against the carrying capacity of the habitat, exactly as in the higher animals. Then, eight or ten thousand years ago, the agricultural revolution began. Food productivity increased by leaps and bounds, and power and wealth accrued to those groups which allowed their populations to multiply, settle villages, and build towns. The old checks on population growth were insensibly for-

gotten and discarded, and none have since taken their place. We should not therefore look to nature to halt the human population explosion: it must depend on our own efforts.

V. C. WYNNE-EDWARDS

[Directly related are the entries ECOLOGY; POPULATION. Other relevant material may be found in EVOLUTION; HOMEOSTASIS; WAR, article on PRIMITIVE WARFARE.]

#### BIBLIOGRAPHY

- JENKINS, DAVID; WATSON, ADAM; and MILLER, G. R. 1963 Population Studies on Red Grouse, *Lagopus lagopus scoticus* (Lath.), in North-east Scotland. *Journal of Animal Ecology* 32:317-376.
- WYNNE-EDWARDS, V. C. 1962 *Animal Dispersion in Relation to Social Behaviour*. Edinburgh and London: Oliver & Boyd; New York: Hafner.
- WYNNE-EDWARDS, V. C. 1963 Intergroup Selection in the Evolution of Social Systems. *Nature* 200:623-626.

### SOCIAL CHANGE

Social change is such a prevalent and often disturbing feature of contemporary life that both the specialist and the layman may be tempted to suppose that it is peculiarly modern. Certainly the extent and rate of change in the modern world are greater than in most past periods, but the static qualities of primitive cultures or archaic civilizations are easily and commonly overstated. Change, at some level and degree, is as characteristic of man's life in organized systems as is orderly persistence.

Indeed, small-scale changes may be an essential component of persistence on a larger scale. For example, given man's biological life cycle, enduring systems, such as kinship organization and government, depend upon orderly patterns of age-specific role performances. Likewise, changing patterns—on a daily, weekly, monthly, or annual cycle—provide a basic and predictable continuity to the patterns of social existence.

Paradoxically, as the rate of social change has accelerated in the real world of experience, the scientific disciplines dealing with man's actions and products have tended to emphasize orderly interdependence and static continuity. The genuine difficulties of dealing with social dynamics are in part responsible for this state of affairs. The relationship between small-scale and large-scale change and the relationship between short-term and long-term change exemplify the many analytical and factual complexities that are involved. These considerations make a formal defi-



inition of social change highly desirable, and we shall therefore attempt one here:

Social change is the significant alteration of social structures (that is, of patterns of social action and interaction), including consequences and manifestations of such structures embodied in norms (rules of conduct), values, and cultural products and symbols.

This definition encompasses small-scale change, such as the gradual development of a leadership role in a small, task-oriented group; cyclical patterns of change, such as the succession of centralization and decentralization in administrative organizations; and revolutionary change, such as the overthrow of a government. It includes short-term changes in employment rates as well as long-term changes in occupational structures; both growth and decline in membership size of social units; continuous processes such as specialization and bureaucratization; and discontinuous processes such as particular technical or social inventions.

### Social change and cultural change

The broad definition given above comprises both what is commonly identified as *social* change, which refers mainly to actual human behavior, and *cultural* change, which refers mainly to culturally meaningful symbols produced by human beings. The emphasis in this discussion will be on the interplay among the complex normative patterns of behavior that we call institutions, since it is these that provide much of the rationale for social control and human activities in general.

This emphasis cuts across conventional distinctions between the "social" and the "cultural" aspects of social systems. Cultural change, it is true, requires social actors as agents, and social change is likely to have cultural counterparts. However, changes in certain cultural subsystems—for example, language, the arts, and perhaps theological or philosophical systems—may be viewed in virtual abstraction from concrete human behavior (see, for instance, Sorokin 1937–1941). Similarly, fluctuations in the fashions of dress may be viewed as "autonomous," although it is also proper to consider such fashions as patterns of appropriate conduct in one sphere of social behavior (Kroeber 1957).

It is true that such modes of abstraction are often mere matters of convenience. For example, the steady specialization of vocabularies in language systems may be treated as a principle of autonomous evolution; it may be related to the expansion of knowledge and to role differentiation in complex social systems. However, there is an

underlying problem that should be made explicit. The degree to which cultural subsystems, such as language, may be traced to structural sources is a question of considerable theoretical importance. Similarly important is the question of the degree to which such cultural subsystems may be translated into guides for social behavior. In human societies the extent of autonomous variability among coexistent features appears to be substantial. Therefore, a multiplicity of principles relating to structural regularities and to significant alterations is necessary for the understanding of order and change.

It follows from the possibility of autonomous variability and from the initial discussion of small-scale social changes that in order to formulate principles of social change we must first of all identify the social structure to which these principles are to be applied. It is also necessary to specify the time period over which change is to be studied and to set up standards for measuring various degrees of change. Until these conditions are fulfilled, we cannot even begin to say what it is that is changing or how much it has changed. There is no singular, sovereign cause for changes in social systems or subsystems. It is true that the scientific quest for simplification has led to the identification of individual variables, such as technological innovation or population growth, that are important enough in themselves; but the result of these inquiries has been the development of special theories of change for specific classes of structures rather than any kind of master theory that embraces all types of factors.

For small-scale social structures in general—face-to-face groups, for example, or formal organizations—we can safely assume that change will originate through such familiar mechanisms as the normative requirements attached to role performances (Moore 1963, p. 50). Large-scale systems, such as whole societies, are less easily studied; but even on present knowledge we can be confident that population growth or decline and the vicissitudes of incorporating infants into the system through socialization will introduce at least some flexibilities and adjustments, if not major structural changes in a definite direction. The probability of both technical and social innovation may be inferred from a universal feature of human societies which can be called the lack of correspondence between the "ideal" and the "actual" in the realm of social values. On the whole, these innovations are likely to be directed toward both "adaptation" to the nonhuman environment, to which adjustment is never perfect,

and social control of the human population. Recognition of both of these elements can be found in the theories that we now propose to examine.

### Changing theoretical interests

The nineteenth-century predecessors of modern sociology were very preoccupied with the dynamics of social change. Although some scholars, such as Frédéric Le Play, attempted to establish canons for systematic description of contemporary social types, the attempt to trace the paths of history was a far more prevalent concern. Often the history attended to was not universal, but limited to the fairly clear antecedents of European civilization. Generally these authors attempted to find order in the succession of civilizations [see *HISTORY, article on THE PHILOSOPHY OF HISTORY*]. The most ambitious of them was probably Auguste Comte, who invented the term "sociology" and propounded a "law of the three stages"—the theological, metaphysical, and positivist—to which all civilization was supposed to conform.

**Evolutionary theories.** The directionality of change, and in particular the increasing complexity and structural differentiation of society, came to be a major tenet of evolutionary theories. Following the impact of Darwin's theories of biological evolution, Herbert Spencer, Lewis Henry Morgan, and others of lesser stature used such Darwinian notions as selective adaptation to account for both the cross-sectional diversity of societies and cultures and the supposedly sequential stages of social organization. By the end of the nineteenth century, evolutionary theory was a dominant factor in social thought, even in the work of writers who were not predominantly evolutionist in outlook. This applies to theorists as diverse as William Graham Sumner, despite his predominant concern with the relativism of all social values, and Émile Durkheim, whose life-long devotion to explaining social phenomena in terms of the balance of an interdependent system has caused him to be identified with what later became the "functionalist" approach to society.

**Marxist theory.** Even Marxism was a variant of evolutionism, particularly in its adherence to the notion of sequential stages of social organization. The Marxists tried to show how social change came about by laying great stress on the interaction of technology with social organization. Indeed, Marxist thought in its crudest form shared with most evolutionary theory a belief that one stage of social organization succeeded another through the operation of forces that were as impersonal as they were inevitable. Marx himself,

however, took fairly full account of the purposive character of social action; he did not rely solely on his theory of revolutionary change. Moreover, his theory was a dynamic one, although it underplayed the independent role of ideas and values; thus his intellectual heirs were never caught up in the extremes of static "functionalism" that later became a dominant theme in anthropological and sociological theory.

**Functionalist theory.** Functionalism is the attempt to explain social phenomena by other social phenomena that are contemporary or quasi-simultaneous. In this respect, it rejects the "quest for origins." Some of its proponents (e.g., Durkheim, Radcliffe-Brown, Malinowski), in eschewing explanation in genetic terms, also tended to suppress all queries about the actual dynamics of change. For them, the demonstration of interdependence between different elements of social structure came to mean the search for self-equilibrating mechanisms in society.

The recent revival of interest in dynamics owes something to all of these precedents. From functionalism, contemporary theory derives not only notions of systemic linkages which *may* be sequential, but also, through the concept of "dysfunction," some notions of tension and incipient change. However, the renewal of concern with analysis of social change probably owes more to the undeniable facts of contemporary life and particularly to the social scientists' resultant involvement in studies of modernization. Although functional systems models have served rather well as predictors of eventual expected changes in elements of social structure held to be related to economic development, the unconfirmed assumption that all fully modern societies have essentially the same kind of social structure can only lead to an unwarranted sociological determinism. In any event, this mode of analysis has produced a kind of "comparative statics," offering a before-and-after view curiously discordant with older evolutionary theories; for the functionalists' standard treatment of modernization dwells on pre-existing heterogeneity that becomes in due course a homogeneity. Nevertheless, functionalist studies of modernization have at least the merit of demonstrating a serious but fortunately remediable weakness in the analytical models employed, namely, that before-and-after comparison diverts attention from the mechanisms of change and entirely suppresses inquiries about actual sequences and timetables.

**Alternative models of societies.** Any theory that views society as a "functional equilibrium system" has the advantage of telling us that certain vari-



ables need to be correlated with certain other variables if the theory is to prove its worth. In any case, some such model of society, whether implicit or explicit, underlies most of the better attested propositions about social action and social structure in contemporary American sociology; indeed, these propositions are American sociology to a great extent, and their influence on other branches of social science is already profound. In spite of the objections already brought against functionalist theory, its prevalence is not necessarily to be deplored. Functionalism, in the hands of a sophisticated theorist, does not preclude attention to deviance, nor does it prejudge the issues of stability and continuity in social change theory. Indeed, to specify "the functional prerequisites of a society" (Aberle et al. 1950) is to adopt an implicitly evolutionary position, for the notion of functional prerequisites not only gives rise to high-level generalizations about the features common to all societies but also, by applying the test of survival, explains the failure of past society by their want of one or more "operating characteristics." The sophisticated functionalist, therefore, views society as a "tension-management system," thus making order itself problematic; he welcomes empirical study of the uncertainties and conflicts that arise in society from disparities between the ideal and the actual.

One of the few viable alternatives to functionalism that have been offered is the "conflict model" of society (Dahrendorf 1957). However, this does not appear to be a comprehensive construction, but only a change of emphasis to discordant elements and interests as a counterpoise to models that stress social integration. Moreover, the tension-management model, as described above, already identifies structural tension points as probable sites of change.

### Problems of detection and measurement

Change takes place in *time*, and time for observation is limited, especially for the single investigator. One method that has been devised for coping with this problem is the "panel study" [see *PANEL STUDIES*]. This is a method that attempts, over relatively brief periods, to record changes in individual attitude and behavior while they are in process. For longer periods, there would appear to be no substitute for historical evidence. Here the social scientist commonly turns to the professional historian; but the historian's interests do not often coincide with those of the sociologist, and it is not surprising if data that would help to answer sociological questions are either passed over or dealt

with only briefly in works of history. Moreover, historians have rarely made use of quantitative techniques of data reporting, and the sociologist interested in constructing actual rates of change may have to do his own history. Finally, there is the undeniable fact that much of the past is irretrievably lost to human knowledge. The assembling of social statistics is very recent in man's history and still very incompletely practiced [see *CENSUS; GOVERNMENT STATISTICS; LABOR FORCE, article on DEFINITIONS AND MEASUREMENT; VITAL STATISTICS*].

**The problem of magnitude.** It was earlier noted that the question "What is changing?" is primary. The identification of elements or constructs for observation and analysis is a theoretical matter. Opinions may properly differ as to what constitutes a "significant" change; there is no legitimate scientific basis for denying social scientists their interest in the tiny alterations that occur in small groups or in the day-to-day developments of a political campaign. Questions of magnitude, however, cannot be avoided if we are to talk about social change in general. How much of a change is to be taken as "significant" and by what criteria of significance?

If the system constructs explicitly or implicitly employed by some functionalists were to be taken seriously, this problem could be solved simply by the strict dichotomization of change as "trivial" or "tragic" (Moore 1963, pp. 20, 70). In its extreme form, the assumption that all parts of a social system are completely interdependent results in the dilemma that change must be viewed either as constituting only variations on a theme within tolerable limits of persistent order ("trivial") or as so altering an essential social component as to have serious repercussions throughout the entire system ("tragic"). However, the problem is not that simple. The structural looseness of social systems is not only evident but empirically verifiable; it is therefore possible that some small-scale changes alter social systems by cumulative degrees. How much change, in what, has what predictable consequences for which specifiable classes of social phenomena?

**A demographic example.** Some examples from demographic behavior may help to clarify the issue. Let us suppose that mortality rates are relatively low and temporally stable and relatively uniform in terms of such standard social categories as residence, ethnicity, occupation, and income. Let us also suppose that birth rates are relatively low and that the modal number of children per family is again relatively uniform across social categories.

Now let us suppose a fairly rapid rise in birth rates, which in familial terms means an increase in the modal number of children from two to three. Such a change will have only minor consequences for the structure of the family, the consequences being limited to the nuances of generational and sibling relations. But the *aggregative* effect of this change on the economy, the density of population, the demand for schools and teachers, and so on, will be major; indeed, if we make allowances for the simplified nature of this example, it can serve as a description of the "baby boom" following World War II. Now let us alter the assumption of the independence of family size with respect to various social categories. Let us suppose, rather, that over the course of a few decades the average size of family changes from a distribution inversely correlated with measures of socioeconomic status to a strongly positive correlation. Children would then become another symbol of affluence, but the effects of the change would be more complex. For example, with regard to mobility opportunities, it would appear that the relative smallness of poor families would favor a more equitable distribution of meager resources for current support and continuing education. Yet the advantages accruing to the children of well-to-do-families would probably persist; indeed, to the degree that inheritance of status continues to operate, the saturation of the upper ranks would preclude substantial upward mobility on the part of children of humble origin. In fact, if the fertility differentials were sharp, only a changed status distribution resembling an inverted pyramid could avoid substantial downward mobility among the well-born.

These partially hypothetical illustrations of demographic trends underscore the importance of looking for consequences of any kind, whether they are direct or indirect, small-scale or large-scale.

**Problems of identification and observation.** The ease with which social change can be identified or observed depends on at least four variables, all of which characterize the dynamic pattern of change. These variables are scale, brevity, repetition, and mensuration.

**Scale.** "Scale" is a somewhat ambiguous term, which may refer to the size or comprehensiveness of the system affected—for example, the political structure of a whole society as compared with a local custom relating to courtship—or to the degree of alteration in the system. As magnitude in the latter sense implies at least crude measurement, the "scale" variable is most appropriately reserved to refer to the comprehensiveness of the system involved. These two variables—the size of the sys-

tem and the degree of alteration—are of course likely to be positively correlated. However, the correlation is not perfect; small changes can occur in inclusive systems and large changes in subsystems.

**Brevity.** Changes that are completed in a relatively short time are likely to be more observable than those that take place over a longer time, even if the latter eventually become substantial in magnitude. Long-term cycles which occur on a large scale may be detected retrospectively; however, they will probably go unnoticed if they are on a small scale and involve only small degrees of variation.

**Repetition.** Relatively brief changes, even if of small magnitude, have a greater chance of being observed if they are repeated frequently—in a cyclical pattern, for example. Repetition, in fact, greatly facilitates explanation and prediction, in that chance antecedents may be more readily distinguished from common and efficacious ones. In this respect, the repetition need not be confined to a single social unit, provided that there is enough similarity among units to allow for comparison. Thus some of the social changes related to industrialization are unique and discontinuous with respect to past trends in areas undergoing modernization, but they are sufficiently repetitive from one society to another (time for these purposes being only an incidental variable) to permit the formulation of general principles.

**Mensuration.** For large-scale transformations in social systems, such as political revolutions or rapid industrialization, the quest for quantities may seem unnecessarily laborious. Yet even here numerical indicators are likely to be useful for comparing cases, and they are particularly necessary if one is going to speak at all precisely about rates of change. In the case of less dramatic social transformations, the perception of change may not be uniform, and reliable measurement may therefore be the only way of settling the issue, as well as of finding at least partial explanation for clearly demonstrated changes. For example, is mental illness increasing in the United States and other modernized societies? If so, by what standardized criteria of identification and diagnosis can the increase be demonstrated? If it is demonstrable, how can cross-sectional differences in rates—for example, by age or income—be converted hypothetically into temporal trends, still using these same indicators? Failure to exercise both quantitative ingenuity and great methodological caution is likely to produce only a spurious explanation for a fictitious trend; such academic exercises unfortunately have been too common.



**Demonstration of causal processes.** It is very rare, in studies of social change, that we come across a classically neat demonstration of a singular cause, producing a singular effect, under finitely specifiable and repeatable conditions. This is partly because of the difficulty or impossibility of fully controlled social experimentation. A further difficulty is the inadequacy of the currently available conceptual and observational tools. Usually the analyst deals with statistical probability distributions and attempts to explain various "dependent variables" so characterized in terms of "independent variables" having the same uncertainties. Statistical techniques have been developed for dealing with these crudities, and others are likely to be invented [see CROSS-SECTION ANALYSIS].

Sequential analysis, however, involves additional problems, for it takes into account not only the degree to which the outcome is determined by one or another measurable antecedent but also the important question of the temporal order of critical variables. Smelser (1962, pp. 12-21) has borrowed from economics the term "value-added process" to refer to this type of logical scheme for organizing the determinant variables. This methodological principle offers the analyst some prospect of moving from the demonstration of necessary conditions to the understanding of sufficient conditions for social change, under favorable circumstances of observation and measurement. The notion of "value-added process" resembles various "stage" theories of change, although stages often represent arbitrarily imposed discontinuities in cumulative trends and thus understate the interaction of variables in the process of transformation.

One of the more complex forms of evolutionary change has been identified by Arnold Feldman and Wilbert E. Moore (1960) as cumulative, retroactive evolution. This term refers to the process by which later "stages" have a continuing or delayed effect on seemingly earlier ones. The mathematical conception of "stochastic processes" is a more accurate designation for this process than the concept of "stages." However, the delineation of stages is a useful mode of characterizing sequences that exhibit fairly clear breaks in rates of change and "emergences" that fundamentally alter the properties of social systems.

### The directions of change

In accordance with the position that only "special" theories of social change are appropriate to the diversity of social phenomena, the varieties of directionality in change will each be illustrated empirically. Although many of these have been

offered in the past as master principles of social dynamics, the eclectic view here espoused has clear advantages in terms of factual confirmation, even if it thereby loses in simplicity and level of generalization.

**Evolutionism and relativism.** The principal mistake of the evolutionary theorists of the latter half of the nineteenth century was overgeneralization. The rectilinear theories of human evolution either took too little account of diversity or explained it improperly. Although Spencer's concept of the "superorganic" offered promise of recognition that human societies might show some portion of the diversity of plant and animal species, he was too bound by the idea of the biological unity of mankind to exploit his fundamental idea. Later theorists went to the opposite extreme by ignoring biological factors and emphasizing sheer unexplained diversity. This extreme relativistic view was no more acceptable than the evolutionists' idea of lining up all of the world's cultures in rank order, with the presumably "simpler" social organization representing a kind of unexplained primitive survival.

**Fluctuations and cycles.** In the analysis of social dynamics, the counterpart of the recognition of simple cross-sectional diversity has been the identification of "fluctuations," that is, variations through time which show no clearly repetitive or cumulative pattern. Fluctuations do indeed occur; for example, consider the chance variability in reproduction and the uncertainties of the intersection of hereditary and environmental factors in personality formation. In fact, the vicissitudes of matching recruits with the normative requirements of enduring positions insure some fluctuations in the exact behavioral counterparts of established structures (Sorokin 1937-1941, vol. 4; compare Moore 1963, pp. 66-68).

As compared to fluctuations, cycles involve more determinate directionality. A simple pendulum swing between extremes (e.g., prosperity and depression, liberalism and conservatism, innovation and accommodation) represents the simplest cyclical pattern. The delineation of multistage cycles is likely to rest on tenuous analogies (e.g., the "life cycles" of civilizations) or to require somewhat arbitrary breaks in variables that on inspection show rather orderly trends.

**The self-equilibrating model.** The notion of "self-equilibrating mechanisms" in functional systems is more readily understood in terms of trendless cycles than it is in terms of a static continuity which is only occasionally disrupted by external disturbances. For example, there is a tendency in

administrative organizations for "indulgent" departures from strict compliance with the rules to grow more extreme, until reversed by a renewed insistence on approximate compliance. This might be characterized as the "cycle of sin and penance." Yet the notion of "self-equilibration" often implies that trends are reversed because of side effects of actions taken by the participants. This mechanistic assumption is not always warranted; the question as to whether the direction of change is deliberate is always a proper one.

A further point about cyclical change may be derived from the example of a small-scale repetitive pattern just used. Cycles may be superimposed on an underlying, cumulative trend. Thus the "return to the rules" in an administrative organization is unlikely to be complete, and through successive repetitions of the sequence there is likely to result a long-term modification of the norms themselves. On a larger scale, the alternation of relatively high and relatively depressed economic indicators in complex economies may represent essentially short-term variations that are consistent with long-term growth as indicated by per capita income or similar measures of economic development.

**Progress as an aspect of growth.** The evolutionary theorists of the nineteenth century were also naive in that they exaggerated the cumulative nature of change and comfortably equated change with "progress." However, to some extent the cumulation of which they spoke was real—a point missed by all their relativistic detractors. For various sectors of any social system, and even for entire systems, there is evidence of steady or even accelerating growth over long periods of time. For instance, the growth in the number of rules in continuing organizations is certainly not at a steady rate if very frequent temporal comparisons are made; however, it is probably very steady over somewhat longer intervals. Likewise, the growth in total human population has been variable over considerably longer periods, and yet cumulative over the entire span of man's earthly tenure. Rates of technical innovation are variable according to time and place, although commonly on a cumulative base. If the invention rate is the unit of observation, its trends may appear nearly cyclical over extensive periods of human history. On the other hand, if the sum total of useful knowledge is taken as the basis of observation, the short-term variations in the rate of addition to stock are likely to appear as very minor fluctuations in the long-term accumulation of reliable knowledge. This is because the growth of knowledge takes place at an

*exponential* rate: the more there is, the faster it increases.

**Measurement of trends.** Conclusions about both the rate and the direction of social change are much affected by the choice of time intervals of observation, including the span between the initial and terminal observations. Constant surveillance of ongoing processes is rare outside the laboratory and nearly impossible for long-range changes. Frequent recording of observations tends to pick up cyclical patterns and essentially meaningless fluctuations. Infrequent recording loses such information and points instead to more enduring trends. This is not to argue for less information, as infrequent observations are subject to grave risks of sampling error in the peculiarities of the observational situation. Skilled analysis can distinguish mere fluctuations, short-term trends, and long-term trends from a "continuous" series. The fact is, however, that in the retrospective analysis of long-term changes the analyst is more likely to detect a "fundamental" trend; thus the finer processes of transformation may be lost to view.

**Mortality and fertility trends.** A further example of great substantive importance may serve to underscore the problems of sorting out short-term and long-term trends. Contemporary evidence and all reasonable inferences from available historical materials indicate fairly wide short-term fluctuations in death and birth rates. But they also indicate two clear longer-term patterns. Over the very long term, human population growth has followed the cumulative, accelerating (*exponential*) pattern. Examined in somewhat more detail, the shorter-term patterns exhibit greater variability in mortality than in fertility in "nonmodern" societies and greater variability in fertility in modern industrialized societies.

Furthermore, the steady decline in mortality, followed, after an interval, by a steady decline in fertility, are trends uniformly accompanying economic development. The trends represent a dynamic pattern that is repetitive, although the pattern seems to operate only once in each economic unit.

This example can be pushed still further with respect to the direction of change. The variation in mortality in "premodern" societies appears to be rather irregular, as it is determined by such "chance" factors as climate and harvests, political vicissitudes, and infectious diseases. On the other hand, the variation of fertility rates in "modernized" countries seems to approximate a cyclical pattern. In societies where contraception is practiced, birth rates are likely to be correlated with *current* economic indicators [see FERTILITY CONTROL].



However, there is more than a slight suggestion that fertility functions not only as a dependent variable but also as an independent variable. Prosperous conditions result in exceptionally high fertility. As these birth cohorts reach maturity they tend to oversupply the prevailing labor demand, which serves to increase unemployment and other unfavorable economic indicators. These deteriorating economic conditions in turn result in small birth cohorts. Note that if this reasoning has further empirical validation (and it fits rather well the fertility behavior of Western countries for several decades following the 1930s) we shall have an example of a "self-disequilibrating" system. The cycles remain "off-phase" because of the intrinsic lag entailed in bringing infants to the age of labor-force participation.

**Cumulative trends.** The ways in which cumulative trends display themselves are numerous. To simple, additive (rectilinear) change and accelerating (exponential) growth should be added one form or another of "logistic" trend. A slow increase accelerates because of an increasing base for combinations to take place, but then it "dampens off" as severely limiting conditions are encountered. The size of organizations and even of entire populations often display this pattern. However, it must be remembered that the parameters or restrictive boundaries are themselves subject to change—for example, by technological or organizational innovation. Moreover, the lower limits, as well as the upper, may serve as restrictive boundaries. For example, the decline of mortality is commonly described as a "reverse J-curve," meaning that early and easy improvements in health conditions are followed by more and more difficult life-saving measures until the limits of biological survival are approached.

### Structural differentiation and social change

The directionality of change can be represented in terms of measures of size, incidence, or occurrence. Moreover, certain *processes* of change may be found to have a reliable and enduring direction. Most notable of these in the theoretical literature is the presumably universal tendency toward specialization or structural differentiation. But the question of why this should be so is commonly left unanswered, the process being taken as given. Durkheim (1893) and others sought a demographic explanation: growing numbers lead to more complex social arrangements, including positional and role differentiation. Extensive evidence does point to a high, but by no means perfect, correlation between size and specialization; however, the reasons

for this correlation are left for the most part unexplained.

In Darwinian evolutionary theory, structural differentiation derives from selective adaptation of organisms to their environment. Since environments differ both cross-sectionally and temporally, the idea of selective adaptation provides a way of accounting both for the observed diversity in structural forms and for continuing change. It is surprising that so little use has been made of this conceptual scheme in the theory of social systems, where it appears equally applicable. In this scheme, population or membership size would serve as a basis for increased potential variability, and differentiation internal to the system would be one mode of adaptation of the entire system to its environmental setting. Criteria of efficiency in optimizing goal achievement—criteria which may be overt and purposive in complex social systems—would then be particular forms of adaptation.

Specialization is not, of course, an absolutely sovereign and irreversible dynamic process. Occasionally, its dangers to systemic cohesion may lead to renewed emphasis on unity and commonality of participating units. However, the probability of continuing specialization in enduring social systems is high. The cumulative growth of reliable knowledge and technique abets the requirement of differentiated competence.

Structural differentiation does imperil systemic cohesion, as just indicated. Mere interdependence, as Durkheim (1893) demonstrated, does not insure "solidarity." Solidarity, or systemic cohesion, also requires an effective system of norms, which in turn will be justified or rationalized by common values. The *mechanisms* of effecting coordination are chiefly two: exchange, whether through relatively impersonal monetary markets or other forms of complementary reciprocities; and administrative authority, a mode of allocating duties and insuring compliance by the exercise of institutionalized power. These mechanisms are not of course mutually exclusive; for example, subordination to authority may be contractual.

A supplementary note on directionality is appropriate. Any of the forms of change we have identified may be associated with ideas of "progress," which is simply change in an evaluatively approved direction. It is of course not itself a scientific principle, but as an item of belief it is possibly a relevant social datum. Progress may be viewed by social participants as the continuation of cumulative growth, as the dampening of cyclical oscillations and return to a steady state, or even the return to a past state which is evaluated as superior

(the concept of "primitivism"). But since social action is partly purposive, the direction of change may well reflect the *desired* direction of change. Thus ideas of progress may be important elements in accounting for observed dynamics.

### The problem of discontinuities

Change is most readily perceived as "orderly" when measured trends display a uniform direction or rate or when cycles are repeated often enough to confirm their character. Even lawful acceleration (Moore 1964) such as the exponential curve, which makes the rate of growth proportional to the extent of the relevant universe at any time, displays an underlying order. From a theoretical point of view, such orderly change clearly requires persistence or repetition in the ambient conditions. From a practical or procedural point of view, orderly change provides sufficient time for the identification of causes and the refinement of observation and measurement.

Analytical difficulties multiply when change is in a significant sense and degree "discontinuous." For example, extremely rapid change, even if its antecedents are clear, may in turn alter the fundamental structure of the underlying change-producing conditions. The "successor system" which emerges can be described as "qualitatively" different. This does not mean that measurement is no longer possible or appropriate; it does mean that new variables and parameters must be taken into account.

Two illustrations from the contemporary world are relevant here. In retrospect, the discovery of nuclear fission and fusion may be viewed simply as an evolutionary step in the pattern of multiplying power utilization (White 1949). But the *implications* of that power for international politics, space travel, and economic production have introduced new dimensions into the potentialities for further change. At the more strictly social level, the extremely rapid spread of the doctrine of economic development has substantially reduced the cultural insularity of tribal societies and archaic agrarian civilizations and thus has increased the theoretical and practical utility of viewing the entire world as a single system.

Closely related to such "emergent" phenomena as the social and political effects of atomic power is another class of events that we propose to call "threshold" phenomena. By this we mean that the cumulative and interactive effect of changes that are analytically separable may result in unprecedented transformation. Thus, although there is no precedent for a *rapid* decline of birth rates in the

course of economic development, several antecedent conditions for a sharp break with the past may be noted. They include the potentiality and actuality of mortality reductions at a much more rapid rate than in the oldest industrial countries; a nearly universal political adherence to rapid economic development, coupled with growing recognition that economic programs are imperiled by extremely fast population growth; and a steadily changing "technology of contraception" that may greatly increase its economic and aesthetic (and perhaps even religious) acceptability.

The concept of the threshold emphasizes sequential rather than factorial analysis of causation. Like Smelser's use of the "value added" concept (Smelser 1962), it represents an attempt to move from necessary to sufficient conditions in the prediction of change, including alterations in direction, rate, or subsequent characteristics of the social system which merit the designation "discontinuous."

**Reversal of trends.** Trends may reverse as well as cumulate to new levels. The upswing in birth rates among industrially advanced countries following World War II illustrates completion of a process of change, after which the number of children per family began to be determined by previously unimportant factors. The historic decline in fertility owed something to marriage rates and ages, but chiefly reflected an extending pattern of contraception. As indicated by the then standard inverse relationship between fertility and socioeconomic status, the leading practitioners of contraception were those families which could best afford children. But in the course of time the attitudes and techniques favorable to family limitation became less and less the exclusive possession of upper-class and middle-class families. Apparently, it was chiefly education, among the various criteria of status, that had the greatest influence on family limitation, and it is also noteworthy that it is education that has exhibited the greatest equality of distribution. When a third or fourth child was no longer putative evidence of ignorance and error, but rather might be interpreted as a choice of family size, birth rates began to take on the characteristic trends and distributions of the markets for "consumer durables."

**Persistence of traditional structures.** Similarly, economic modernization is never "complete." Its initial, discontinuous impact on traditionally organized societies may be followed by partial restoration of structures and forms which were never completely destroyed. For example, the weakening of extended kinship systems as a consequence of



geographic and social mobility may be followed by partially restored, though discretionary, communication and reciprocities among kinsmen. Likewise, stable exchange or employment relations are likely to lose some of their nominal austerity and impersonality. Note that this is not to say that the *status quo ante* is fully restored or that nothing significant has changed, but only that the conception of an unusually disruptive "transitional period" has some merit.

**Revolution as discontinuous change.** Fundamental "social revolutions" represent an extreme form of discontinuous change. Although social scientists have been at some pains to emphasize the partial restoration that occurs even after major revolutions, by definition revolutions involve fundamental changes in institutional structure and distribution of power. Polarization (for instance, of social classes of interest groups) is of course one antecedent of revolutionary change; however, polarization in turn requires explanation. Under the conditions of political centralization, substantial urban agglomeration, and fairly effective communication, polarization is likely to occur when there is an absolute or relative decline in economic well-being or political rights among those already largely excluded from positions of affluence and power. The decline, of course, must be fairly severe, sharply differential in its impact, and numerically extensive. Although this form of change is implicitly quantifiable, very little quantitative analysis of the antecedents of revolution has been attempted. Polarization, for example, is typical in the modernized sector of societies which are in early stages of industrialization. Yet in the majority of historical cases that sector has been too small to provide a clear revolutionary potential, and as it has expanded it has also reduced its polarity. How extensive and rapid must "early" modernization be in order for it to produce social revolution? The apparently faster change in aspirations than in the means for their fulfillment in contemporary developing areas invites the speculation that the incidence of revolutions may increase in the future. Refinement of that speculation into a responsible probability distribution awaits the collection of comparative data, both cross-sectional and temporal.

### Forecasting the future

Theories of social change have little claim to scientific merit unless they are formulated as verifiable predictive propositions concerning the relations among variables. However, in the "real world"

the interaction of factors is very complex, and it may be difficult to replicate or control the conditions under which the causal variables are sequentially related. Thus, generally speaking, small-scale and short-term changes are easier to predict than their opposites. Science cannot predict the unique event with absolute reliability; it can only assign probabilities to individual occurrences in a class of events. In this respect, the task of the historian, whose data are in principle "all in," is substantially easier than that of the sociological prophet, who may not foresee crucial intervening events or even be able to assign correct values to available information.

Although the precise shape of the future is hidden by a haze, there are several bases for at least rough and partial prediction. The first of these is the simple *persistence* of past and present conditions. Fortunately, for much of social conduct we can make this assumption; otherwise we would not know how to act or what to expect from day to day. Customs, organizations, and values may be expected to survive the pressures of other changes over long periods of time. For example, we may predict with fair confidence that the United States will continue to have a constitutional government; although it may experience some shifts in political influence and even in the relative powers of the executive, the legislature, and the judiciary, it is not likely to become a one-party regime or adopt a strictly executive mode of government. Moreover, the expansion of the size and functions of government does not lead us to expect the end of pluralism or any substantial reduction in the large measures of individual and private-group choice, decision, and competition which are characteristic of the "open society." Some American traits, such as pragmatism in attacking problems and a kind of rational, secular view of most social institutions, look like hardy survivors from the past which will continue into the future.

A second basis for prediction is to be found in the *continuation of orderly trends*. In addition to our earlier discussion, some other probable trends may be noted. In the United States, the rate of urbanization and suburbanization has been increasing, but changes in the rate have been fairly orderly over three decades. The average age of marriage has been decreasing, but at a slowing rate. The proportion of secondary school graduates has been increasing, as has the proportion going on to college. Women are entering the labor force at a gradually rising rate, particularly at ages over 45 and under 60. The amount of private saving increases as the

economy grows and income increases, but the proportion of money saved remains remarkably constant. Some of these trends have a decade or two of history behind them and some much longer periods. With the extension and improvement of social accounting, the examination of trends may become an increasingly important basis for forecasting.

Some destinations are known because the route has been traveled before. In addition to small-scale repetitive patterns, the contemporary world offers examples of large-scale recapitulated experience. In the course of economic modernization, precise replication of "Western" rates and sequences of change will not occur often. However, on the basis of what is known about "Western" experience we can go rather far in predicting the course of social change in major parts of the social organization and value systems of societies which are now seeking to become part of the modern world.

Another component of forecasting is the great and growing importance of planning, which is of special significance in industrial countries and those attempting to become industrialized. A remarkable amount of energy and other scarce resources is spent on forecasting autonomous trends and calculating intermediate adaptations and on deliberately implementing future goals. Although accidental and "mindless" change still occurs, social change is increasingly both organized and institutionalized. The future is partially predictable because it will resemble in part what it is now intended to be.

WILBERT E. MOORE

[Directly related are the entries CREATIVITY, article on SOCIAL ASPECTS; CULTURE, article on CULTURE CHANGE; EVOLUTION, articles on CULTURAL EVOLUTION and SOCIAL EVOLUTION; FUNCTIONAL ANALYSIS; INDUSTRIALIZATION; MODERNIZATION; REVOLUTION; SYSTEMS ANALYSIS, article on SOCIAL SYSTEMS. Other relevant material may be found in AUTOMATION; CITY; CONFLICT; COOPERATION; ECONOMY AND SOCIETY; FASHION; FERTILITY; FOOD, article on WORLD PROBLEMS; INNOVATION; MORTALITY; PLANNING, ECONOMIC; PLANNING, SOCIAL; POPULATION, articles on POPULATION THEORIES and POPULATION GROWTH; PREDICTION AND FORECASTING, ECONOMIC; SCIENCE, article on THE SOCIOLOGY OF SCIENCE; TRANSPORTATION; and in the biographies of DARWIN; DURKHEIM; LE PLAY; MARX; MORGAN; LEWIS HENRY; OGBURN; SOROKIN; SPENCER; SUMNER; WEBER, MAX.]

#### BIBLIOGRAPHY

ABERLE, DAVID F. et al. 1950 The Functional Prerequisites of a Society. *Ethics* 60:100-111.

DAHRENDORF, RALF (1957) 1959 *Class and Class Conflict in Industrial Society*. Rev. & enl. ed. Stanford Univ. Press. → First published as *Soziale Klassen und Klassenkonflikt in der industriellen Gesellschaft*.

DURKHEIM, ÉMILE (1893) 1960 *The Division of Labor in Society*. 2d ed. Glencoe, Ill.: Free Press. → First published as *De la division du travail social*.

*Economic Development and Cultural Change*. → A quarterly journal, with occasional monographic supplements, published since 1952 by the University of Chicago Press. Especially valuable for theoretical and empirical studies of modernization.

ETZIONI, AMITAI, and ETZIONI, EVA (editors) 1964 *Social Change: Sources, Patterns, and Consequences*. New York: Basic Books.

FELDMAN, ARNOLD S.; and MOORE, WILBERT E. 1960 Moot Points in the Theory. Pages 360-368 in Wilbert E. Moore and Arnold S. Feldman (editors), *Labor Commitment and Social Change in Developing Areas*. New York: Social Science Research Council.

HAGEN, EVERETT E. 1962 *On the Theory of Social Change*. Homewood, Ill.: Dorsey.

KROEBER, A. L. 1957 *Style and Civilizations*. Ithaca, N.Y.: Cornell Univ. Press.

MACIVER, ROBERT M.; and PAGE, CHARLES H. (1949) 1961 *Society*. New York: Holt. → Book 3 contains an extensive treatment of social change.

MOORE, WILBERT E. 1963 *Social Change*. Englewood Cliffs, N.J.: Prentice-Hall.

MOORE, WILBERT E. 1964 Predicting Discontinuities in Social Change. *American Sociological Review* 29: 331-338.

OGBURN, WILLIAM F. (1922) 1950 *Social Change With Respect to Culture and Original Nature*. New enl. ed. New York: Viking.

SAHLINS, MARSHALL D.; and SERVICE, E. R. (editors) 1960 *Evolution and Culture*. Ann Arbor: Univ. of Michigan Press.

SMELSER, NEIL J. (1962) 1963 *Theory of Collective Behavior*. New York: Free Press.

SOROKIN, PITIRIM A. (1937-1941) 1962 *Social and Cultural Dynamics*. 4 vols. Totowa, N.J.: Bedminster Press. → Volume 1: *Fluctuation of Forms of Art*. Volume 2: *Fluctuation of Systems of Truth, Ethics, and Law*. Volume 3: *Fluctuation of Social Relationships, War, and Revolution*. Volume 4: *Basic Problems, Principles, and Methods*.

SPENCER, HERBERT (1862) 1958 *First Principles*. New York: DeWitt Revolving Fund. → A general, classic formulation of evolutionary theory from the inorganic, through the organic, to the "superorganic" (sociocultural).

VRIES, EGERT DE 1961 *Man in Rapid Social Change*. Garden City, N.Y.: Doubleday.

WHITE, LESLIE 1949 *The Science of Culture: A Study of Man and Civilization*. New York: Farrar, Straus. → A paperback edition was published in 1958 by Grove Press.

ZOLLSCHAN, GEORGE K.; and HIRSCH, WALTER (editors) 1964 *Explorations in Social Change*. Boston: Houghton Mifflin.

#### SOCIAL CLASS

See under STRATIFICATION, SOCIAL.



## SOCIAL COHESION

See COHESION, SOCIAL.

## SOCIAL CONTRACT

"The social contract" is the term applied, by a long-standing consensus among students of politics, to the political theories of the most famous and influential thinkers of the period reaching from the mid-seventeenth century to the mid-eighteenth century: Thomas Hobbes, 1588-1679; John Locke, 1632-1704; and Jean Jacques Rousseau, 1712-1778. Scholarly tradition has associated the three because of their alleged common insistence that society originated in a contract, compact, or agreement, explicit or tacit, to which each individual concerned consented and, so, removed himself from the "state of nature" and helped set in motion a regime of government under laws, of impartially administered justice, and of civic morality. Recent scholarship, however, has tended to fix attention on other and perhaps more significant reasons for thinking of the "contractarians" as a school, or movement.

## The social contract school

Hobbes, Locke, and Rousseau were all concerned with emphasizing the contract as an explanation of the *nature* rather than of the *origin* of society—that is, to get across the idea that whatever the origin of society, the relation between the individual members of society and their obligation to obey society's government and laws are fundamentally contractual even in the absence of a formal agreement. Put otherwise, none of the three committed himself unambiguously to the "historicity" of a freely negotiated contract among men in the act of emerging from a state of nature. Hobbes, for example, was willing for the subscribers to his "contract" to be bludgeoned into participation by any strong man capable of forcing their compliance. Locke was willing for the participants to be nursed into compliance by a "godlike" king. Rousseau placed great stress on the role of a "founder," or legislator, in midwifing the contract into existence over a long period of time.

The influence of Machiavelli. All three, if not avowed pupils of Niccolò Machiavelli (who was in disrepute during most of the period in question), at least wrote as if they had been profoundly influenced by Machiavelli. Each was less committed to the contract as explaining the origin or even the nature of society than to certain other ideas which Machiavelli first enunciated: (1) Man and society

are not coeval (as earlier thinkers had asserted); rather, society, like government, law, justice, and morality, is a human artifact, which man is free to alter or dispose of as he sees fit. (2) "Natural" man—that is, man as he must have been prior to the founding of society—is not (as the earlier tradition had held) social or political "by nature"; rather, man first participates in society and continues to participate in it out of one form or another of "fear" or "terror" at the thought of the dire things that will happen to him if he does not participate; man "becomes" social or political only by virtue of the qualities that society superimposes upon his nature through a long process of habituation (man, as Machiavelli put it, is neither good nor bad by nature, but merely malleable). (3) The proper concern of political philosophy is not man's "perfection" or "end"—not, in Rousseau's phrase, "men as they might be"—but men as they are; political philosophy must not fritter away its energies on what Machiavelli called imaginary utopias but must treat of the building of actual societies capable of ministering to the needs and wants of men as we actually know them. (4) There exists no natural or divine law that imposes on men anywhere and everywhere "perfect" duties toward one another and toward society itself, of which their "rights" are derivative. Here, indeed, the contractarians went further than Machiavelli, who had not concerned himself with rights at all, and held that the one firm statement we can make about "rights" is that each man is born with a "right" to "preserve himself" and to choose freely the "means" to his self-preservation.

In summary: Their common dedication to these ideas—all "revolutionary" vis-à-vis the "great tradition" and all rooted in Machiavelli—far more than their use of a "social contract," binds Hobbes, Locke, and Rousseau together into a school. [See MACHIAVELLI.] Viewed in this context and irrespective of whether or not they wrote under Machiavelli's direct influence, the three together constitute a "bridge," moored at one end on Machiavelli, that stretches from the political philosophy of the "great tradition" of the West to the predominant political teachings of the present day. Each, in his own way, contributed to the accomplishment of Machiavelli's avowed purpose: to emancipate mankind from the "bad" tradition inherited from Plato, Aristotle, Cicero, Augustine, and Aquinas.

The right of self-preservation. Hobbes, Locke, and Rousseau built their "models" of political society on the "right of self-preservation" and thereby gave to the word "right" a new and revolutionary meaning. The great tradition, insofar as it had

spoken of "rights" at all (both natural and divine law had emphasized duties, not rights), had conceived of every right as encrusted in a complex, or system, of duties and rights. Its rights had been, therefore, "correlative" to duties, in the twofold sense that, first, a man divested himself of his rights if he failed to perform their correlative duties and, second, that "my" right had been understood as carrying with it a correlative duty on "your" part to respect that right. Not so with the right of self-preservation, which the contract theorists attributed to every man at birth; it carried with it no correlative duties, nor were a man's fellows under any obligation to permit him to exercise it.

One might say that the contract theorists divested the word "right" of any moral connotations whatever: Exercising my right of self-preservation, I enter society, or remain in it, because it seems, from a strictly "selfish" point of view, the best "means" at my disposal for preserving my life (or, in Locke's phrase, my "life, liberty, and estate") and not because of any duty toward my fellow men. [See DUTY.]

**Challenges and contributions.** The four major burdens upon each contract theorist, their handling of which marks the significant differences between them qua contract theorists, were (a) to present a picture of natural, "presocial" man (for the purpose of which they used the device of the "state of nature"), which would render plausible the idea of his entering the "state of society" at all; (b) to show that such and such a "state of society," characterized by such and such a "contract," is a transaction that a rational man, armed only with his right of self-preservation, would choose as the best available means for preserving himself; (c) to account for the obligation on the part of the participants in the contract to keep the "promise" they have allegedly made; and (d) to show why persons born and reared in society subsequent to the negotiation of the contract should be understood to have "consented" to the contract's terms.

None of the three great contract theorists was able, in the judgment of subsequent scholarship, to meet these four challenges successfully, which perhaps explains the fact that the contract device disappeared from political philosophy soon after Rousseau.

However, other ideas that derive either from the school as a whole or from one or more of its members have deeply influenced both political events and political thought throughout the intervening period: first, the notion that no society, government, law, or rule of "morality" is legitimate unless

it rests, directly or indirectly, on the consent of the individuals concerned, though not necessarily consent to a contract; second, that the proper concern of political science is the political behavior of individuals and groups of individuals; third, that societies, governments, laws, and notions of right and wrong or of just and unjust are to be judged by the recognition and protection they provide for "inalienable" individual rights, conceived as inherent in all human beings in all times and places; fourth, that one of these rights is the right to live under a democratic government, that is, government subject to popular control understood as involving, in the absence of unanimity, control by the majority; and fifth, that, in the sense that should be decisive for politics, all men are born equal and that one major purpose of government, therefore, should be to promote equality.

### Antecedents of the social contract

The Hobbesian contract marked a turning point—and in some respects a genuine innovation—in the history of political philosophy. But Hobbes possessed a vast knowledge both of history and of traditional political philosophy, and he, along with Locke and Rousseau, is sometimes said to have been indebted for all or some of his basic ideas to one or another of several sources.

**The Greek "conventionalists."** We know, chiefly from passing references in the writings of Plato and Aristotle, that classical political philosophy itself arose in opposition to a long-dominant school of political thought in which the idea of convention played a role comparable to that of the compact in Hobbes. The contentions of that school were that the Greek city-state, the only form of society the Greeks knew, was a product not of nature but of the kind of tacit, constantly evolving, mutual understandings by which, for example, we today think of languages as originating and developing; that this is equally true of government, of law, and of notions of good and bad and just and unjust; and that it is equally true also of the citizens' habit of obeying the laws of his city.

Among the names of the pre-Socratic conventionalists (of whose utterances we possess only fragments), the most prominent is that of Heraclitus, 575–480 B.C., but some of Plato's and Aristotle's archenemies, the Sophists, defended the conventionalist position. We may say of it, in general, that it is very close to the modern social contract theories in its denial, explicit or tacit, of what were to become the central propositions of classical political philosophy but very remote from those theories in its assertions; conventions, as understood



by the conventionalists, carried with them no obligation and certainly not a contractual obligation. But there is this common ground between the conventionalists and the contractarians: they agree that society, justice, and law can have no superhuman or transcendent source and are therefore "man-made."

*Glaucon's statement.* In the beginning, says Glaucon in Book II of Plato's *Republic*, each man thought it good to inflict injustice upon others and had to suffer injustice at the hands of others; all men behaved accordingly. In due course, however, men came to consider the resulting state of affairs intolerable and proceeded to agree to covenants and laws that obliged them to behave less rapaciously. Such, Glaucon continues, is the origin of justice, which is merely what the agreed-upon laws command, and men accept those laws because it is in their interest to do so. Glaucon, we notice, makes the shift from "convention" to "covenant"; he stresses self-interest as the central motivation in political behavior; and he posits (since, on his showing, there must have been a first covenant) an epoch in the past when men were "presocial," "prelegal," and "premoral." All three of these ideas, which were at most present by remote implication in the conventionalists as such, clearly do anticipate the position of the later contract philosophers. (Of course, Plato's purpose in presenting Glaucon's statement is to discredit the position.)

*The Biblical covenants.* The Old Testament recounts numerous instances of "covenants" between Jehovah and particular individuals (e.g., Abraham) and between Jehovah and the people of Israel. Although they perhaps helped to familiarize subsequent generations with the idea of binding agreements or "contracts," they may be dismissed out of hand as sources for the modern contractarians. They were, in the nature of the case, agreements between Jehovah and an already existing society; the "law" to which they subjected the people of that society pre-existed it and was allegedly not of human origin. (But cf. Gough 1936.)

*The medieval social contractarians.* History records numerous examples (the earliest of which probably was the Acts of the Fourth Council of Toledo in 633) of contracts between kings and peoples which limited the power of the king by declaring his subjects' allegiance dependent upon his recognizing certain of their rights, upon his making new laws only after consultation with the wise men of the realm and with their consent, and upon his promising, often in a coronation oath, to rule justly and in accordance with divine law and

the laws of the realm. Since these contracts did not *create* societies (one of the parties to the contracts was a people that already existed) and since they presupposed justice, law, and morality as goods of transcendent origin rather than human artifacts, their relevance here consists solely in the fact that they did keep alive the idea of contracts as a possible source of political obligation. So, too, with the vast body of medieval literature, much of it polemical, devoted to speculation as to the proper form and content of such contracts. (None of the later contractarians conceived of the relation between the people and the government, the ruled and the rulers, in contractual terms.)

Only toward the very end of the medieval period did political philosophers again fix attention on the way society—or particular societies—originated. They posited a state of affairs that existed before political society was organized and pointed to a contract or agreement among individuals as the most probable explanation of the transition from the one condition to the other.

The Spanish writer Salmonius, for example, argued in a book dated 1544 that both political society and law arise from mutual pledges between individuals who "come together" for these purposes. The Spanish Jesuit Juan Mariana, writing in 1599, described a prepolitical age when men lived as individual nomads, without positive law or government. These came into existence through a mutually binding "compact of society" in which, with an eye especially to the defenselessness of infants and children, they recognize the necessity of common measures of protection against acts of violence and set over themselves, as ruler, a man conspicuous for his justness and uprightness. Mariana's Protestant contemporary George Buchanan put forward a similar theory in Scotland in 1579. In all three cases, we are tempted to conclude that we stand in the presence of ideas that are indeed first cousins to those of the modern contractarians. This is true, however, only in the sense that the late medieval contractarians anticipated to some extent the vocabulary of Hobbes and his successors, and, by speaking in terms of contracts between persons, also anticipated to some extent the emphasis on individual consent.

But a vast chasm yawns between Salmonius, Mariana, and Buchanan, on the one hand, and Hobbes, Locke, and Rousseau, on the other. The former, though they posited "prepolitical" men, conceived of them as possessing duties and rights under divine and natural law; far from insisting upon the "prepolitical" condition as "natural" in

contrast to the political condition, which is a human artifact, they deemed the political condition the more natural of the two because it is necessary for man's perfection and because ultimately it is the handiwork of God. The motives they attributed to men for "coming together" are by no means exclusively selfish and utilitarian (they include self-protection and self-preservation, but not as the supreme motive). Finally, they had no difficulty in explaining the binding character of the contract, since both divine and natural law enjoined the keeping of promises. When, moreover, the Roman Catholic writers of the medieval period spoke of political society as contractual in origin, the point they were making obliquely is that political society is *merely* contractual in origin and therefore inferior to the church, which they believed to have come into existence through a direct act of God.

### Contributions to democratic theory

Hobbes, Locke, and Rousseau have much in common besides their use of the compact, or contract, between individuals in a "state of nature" as a device through which to communicate their major teachings about politics. There are, of course, also great differences among the three, some of them topics of continuing controversy among scholars of political theory; this controversy has become enormously complicated. Each of the three philosophers wrote several works on politics, in which apparent inconsistencies, contradictions, and shifts in position on particular issues can be found, so that no scholarly consensus exists as to precisely what any of them taught. Each, in large part on the basis of testimony from self-proclaimed "disciples," has come to be identified with this or that subsequent political event or movement (Hobbes with modern authoritarianism and dictatorship, Locke with constitutional democracy, Rousseau with the French Revolution and "absolute" majority rule), so that the controversy in question has to some extent been affected and sometimes exacerbated by the political sympathies of the commentators. Finally, each of the three—at least according to recent scholarship—expressed himself on occasion with something less than complete candor, so that there is now dispute as to where we are to look, even in one and the same book by one of the three, for the passages in which the author expressed his "true" intentions, as distinguished from those allegedly included for the precise purpose of throwing the general reader "off the scent" or enlisting the reader's prejudices on behalf of the author's position. If, however, we

fix attention exclusively on each author's definitive work dealing with the contract and related topics (for Hobbes, *Leviathan*; for Locke, *The Second Treatise on Civil Government*; for Rousseau, *The Social Contract*) and think of the three together as a school and as a bridge leading from traditional political philosophy to modern political science and if we concentrate on the concept of contract itself, as it evolved from Hobbes to Locke to Rousseau, some conclusions are warranted about the similarities and differences between them and about their contributions to the emergence of present-day democratic theory.

**Individual rights.** For each of the three authors the individual's right of self-preservation and of choosing the means thereto is that which authorizes the individual in the state of nature to enter into an agreement to form a political society and, so, legitimates the contract. For Hobbes and Locke, however, that right was a "natural" right, which belongs to the individual in the very "nature of things"; both regarded the proposition that asserts the right as requiring demonstration, which for each of them took the form: People as we know them, as given to us by nature, act *as if* they possessed such a right; therefore they do possess it. The mature Rousseau of the *Social Contract* attempted no such demonstration, preferring to treat the right to preserve oneself as axiomatic and, above all, as not a natural right. To Hobbes, therefore, belongs the credit for having taken one great step toward contemporary democratic theory by having made of the problem of the right or rights of individuals one of the central problems of modern political philosophy; to Rousseau belongs the credit for having, in anticipation of contemporary democratic theory, freed the defenders of individual rights from the necessity of rooting them in "nature."

**The law of nature.** Hobbes and Locke both appealed to a "law of nature," which for Hobbes defined the minimal rules that, in the very nature of things, men must agree to observe if they are to constitute a society that will really protect them against the dangers of the "state of nature." For Locke the law of nature defined men's rights and duties in the state of nature and was, therefore, the ultimate source of the individual's right, once within civil society, to protection of his "life, liberty, and estate." Both, in any case, revolutionized traditional natural-law teachings, by giving to the law of nature a new meaning and a new content. [See NATURAL LAW.] However, it was the Rousseau of the *Social Contract* who took the drastic step in the



direction of contemporary democratic theory: he jettisoned the "law of nature" altogether and argued that there is no objective standard that the positive law of a particular society should seek to approximate.

**Authority and legitimacy.** Hobbes's contract commits the individual to permanent and irrevocable membership in a political society, whose first and only task, once it is constituted, is to name a sovereign (either a single man, or an oligarchy, or a democratic assembly) who is empowered to make laws, decide disputes, lay down principles of right and wrong, and distinguish between religious truth and heresy; once the sovereign has been named, the citizen owes him absolute obedience in return for that protection against domestic law violators and foreign enemies which allegedly makes of civil society a "good bargain" for the citizen by comparison with the "violent death" that is his well-nigh certain fate in the state of nature. The effect of the contract is to make the will of the sovereign "representative" of the citizen's own will; the sovereign disposes of the entire force of society for effectuating his will, and the citizen, because that is what he has contracted to do, wills every act of the sovereign that does not involve direct sacrifice of his own life. The sovereign, not himself a party to the contract, cannot "rightfully" be called to account by his subjects for the manner in which he performs his task. The individual, although he comes into the contract by virtue of his right of self-preservation, enjoys within society only such rights as the sovereign wills him to have. The Hobbesian model anticipates modern democratic theory only in the sense that it makes of the "people" the remote source of all governmental authority.

In Locke's conception of the contract the individual also assumes a permanent and irrevocable obligation to obey the legislative that the society appoints in its first act following its institution, though only provided that that legislative act for the "public good."

Some commentators, however, have construed that proviso as guaranteeing to the individual rights that the legislative must not invade or deny, upon pain of exceeding its rightful authority (so that, according to these commentators, political power under the Lockean contract is "limited"). Other commentators have dismissed the proviso as operationally insignificant, since the contract neither specifies the rights in question nor envisages machinery through which the individual might assert them (so that, according to these commentators, Locke's legislative is, from the

standpoint of the individual, no more "limited" and, thus, no less absolute than Hobbes's sovereign).

The decisive difference between the Hobbesian contract and the Lockean is that under the latter the "people," when it finds that the government is violating its "trust"—that is, the obligation to act only for the public good—may rightfully resist the government's authority and overthrow it, though the people must proceed at once to install a new legislative, whose laws the individual citizen, still under the "original" contract, is obligated to obey. If we deem the authority of Locke's legislative to be a limited one, without power to deny or invade certain individual rights, we may say that Locke's contract points forward to the bills-of-rights emphasis in contemporary democratic theory.

Rousseau's contract differs radically from Locke's in two decisive respects. Rousseau's individual, when he enters the contract, cedes all of his rights, including his property, to the community, with the clear understanding that he is henceforth to enjoy only such rights as the "general will" of the community vouchsafes to him and is to perform all duties that that general will imposes on him. Rousseau's contract, on the other hand, is not permanent and irrevocable: the "general will"—that is, the body of the citizens in their legislative capacity—must legislate only laws which are general in their purpose (that is, addressed to the common good of the society's members) and general in their application (that is, of such character as to extend the same rights to, and impose the same duties on, each citizen). When the citizen finds himself in the presence of a law that does not meet these requirements, the contract has been violated and its obligation lapses. Rousseau said nothing about the situation that would supervene on such a violation of the contract; but it is certainly Rousseau's "model," not Hobbes's or Locke's, that points forward to the quest in contemporary democratic theory for a legislative process involving procedural guarantees, whose observance the citizens may demand with the threat of withdrawing their obedience. In other words, Rousseau's conception foreshadows the constitutional emphasis in contemporary democratic theory.

**The problem of consent.** As a corollary to the right of self-preservation, Hobbes, Locke, and Rousseau held that a man can be rightfully "bound" only by his own consent. All three encountered great difficulty in explaining why the contract to which the original contractors consent should be binding upon their descendants. Hobbes and Locke "papered over" this difficulty by asserting that the descendants give tacit consent by remaining within

the community and accepting its protection. Neither Hobbes nor Locke, viewed from the standpoint of contemporary democratic theory, made any serious attempt to carry the principle that a man can be bound only by his own consent over into the "model" of political society.

Here, once again, Rousseau broke sharply with his predecessors and sought in two ways to legitimate the laws of his society by the continuing consents, individually given, of the citizens: first, by stipulating in the contract that each citizen shall be required, on coming of age, to opt for consenting to the existing institutions or for withdrawing from the society, and, second, by requiring that no citizen be formally excluded from the deliberations and votes that produce expressions of the "general will." In both these respects, Rousseau brought us very close to two of the major themes of contemporary democratic theory: the emphasis on political equality and the stress on active participation by the citizens in the political process as an indispensable condition for "government by consent."

WILLMOORE KENDALL

[See also the biographies of HOBBS; LOCKE; ROUSSEAU. Directly related are the entries DEMOCRACY; MAJORITY RULE; NATURAL LAW; NATURAL RIGHTS. Other relevant material may be found in POLITICAL THEORY.]

#### BIBLIOGRAPHY

- ALLEN, JOHN W. (1928) 1957 *A History of Political Thought in the Sixteenth Century*. 3d ed. London: Methuen.
- BUCHANAN, GEORGE (1579) 1581 *De iure regni apud Scotos: Dialogus*. 3d ed. Edinburgh: Apud Johannem Rosseum. → Available from University Microfilms.
- GIERKE, OTTO VON (1881) 1958 *Political Theories of the Middle Age*. Cambridge Univ. Press. → First published as "Die publicistischen Lehren des Mittelalters," a section of Volume 3 of Gierke's *Das deutsche Genossenschaftsrecht*. Translated with a famous introduction by Frederic William Maitland.
- GIERKE, OTTO VON (1913) 1934 *Natural Law and the Theory of Society: 1500 to 1800*. Translated with an introduction by Ernest Barker. 2 vols. Cambridge Univ. Press. → A translation of five subsections of Volume 4 of *Das deutsche Genossenschaftsrecht*. A paperback edition was published in 1957 by Beacon.
- GOUGH, JOHN W. (1936) 1957 *The Social Contract: A Critical Study of Its Development*. Oxford: Clarendon.
- HOBBS, THOMAS (1642) 1949 *De cive: Or, The Citizen*. Edited with an introduction by Sterling P. Lamprecht. New York: Appleton. → First published in Latin.
- HOBBS, THOMAS (1650) 1928 *The Elements of Law: Natural and Political*. Edited by Ferdinand Tönnies. Cambridge Univ. Press.
- HOBBS, THOMAS (1651) 1950 *Leviathan*. With an introduction by A. D. Lindsay. New York: Dutton.
- HUME, DAVID (1741-1742) 1912 *Essays Moral, Political and Literary*. 2 vols. Edited by T. H. Green and

T. H. Grose. New York and London: Longmans. → First published as *Essays Moral and Political* and changed to this title in the 1758 edition.

- JOUVENEL, BERTRAND DE 1947 *Essai sur la politique de Rousseau*. Pages 15-132 in Jean Jacques Rousseau, *Du contrat social*. A critical edition. Geneva: Cheval Ailé.
- KENDALL, WILLMOORE (1941) 1959 *John Locke and the Doctrine of Majority-rule*. Urbana: Univ. of Illinois Press.
- LOCKE, JOHN (1690) 1964 *The Second Treatise of Government: An Essay Concerning the True Original, Extent, and End of Civil Government*. Pages 283-446 in John Locke, *Two Treatises of Government*. Cambridge Univ. Press.
- MCCOY, CHARLES N. R. 1963 *The Modernized Theory of Natural Law and the Enlightenment*. Pages 187-221 in Charles N. R. McCoy, *The Structure of Political Thought: A Study in the History of Political Ideas*. New York: McGraw-Hill.
- MARIANA, JUAN (1599) 1948 *The King and the Education of the King*. Washington: Country Dollar Press. → First published in Latin.
- ROUSSEAU, JEAN JACQUES (1762) 1962 *The Social Contract*. London: Dent; New York: Oxford Univ. Press. → First published in French.
- SALAMONI, MARIO (1544) 1955 *De principatu, libros septem nec non orationes ad priores Florentinos*. Recognovit Marius d'Addio. Rome, Università, Istituto di Diritto Pubblico e di Dottrina dello Stato, Pubblicazioni, Series 4, No. 5. Milan (Italy): Giuffrè. → Salamoni is also known by his Latin name Marius Salmonius.
- STRAUSS, LEO (1953) 1959 *Natural Right and History*. Univ. of Chicago Press.
- STRAUSS, LEO, and CROPSEY, JOSEPH (editors) 1963 *History of Political Philosophy*. Chicago: Rand McNally. → See especially pages 354-378, "Thomas Hobbes," by Laurence Berns; pages 413-432, "Benedict Spinoza," by Stanley Rosen; pages 433-468, "John Locke," by Robert Goldwin; pages 491-513, "David Hume," by Robert Hill; and pages 514-535, "Jean Jacques Rousseau," by Allan Bloom.
- TUSSMAN, JOSEPH 1960 *Obligation and the Body Politic*. New York: Oxford Univ. Press.

## SOCIAL CONTROL

### I. THE CONCEPT

Jesse R. Pitts

### II. ORGANIZATIONAL ASPECTS

Amitai Etzioni

#### I

#### THE CONCEPT

Much of the impetus for the development and use of the concept of social control comes from the sociological adaptation of the Darwinian tradition. But there, however, the major dichotomy was between organism and nature; for the various theories of social control, it has been that between individual and society. It is assumed in these theories that society has to control the animal nature of man: if order is to be established and main-



tained, man's tendency to pursue his self-interest to the point of a war of all against all must be limited through learning or selection, or both. Emergence of the concept of social control thus indicated a waning of the utilitarian concept of the natural harmony of self-interests.

Although social control is essentially an American term, it has its functional equivalents in European sociology. Durkheim saw the *conscience collective* as constraining men, with a power directly proportional to the intensity of the interaction around its specific representations, to behave in certain ways regardless of their own selfish interests. Thus a major function of social institutions such as the family, marriage, and religious cults was to increase the constraining power of the *conscience collective*; institutions were essentially agencies of social control. Implicit in Durkheim's account of society was a concept of deviance as caused by "animal spirits": without social constraint men were prone to violence and to selfish calculations, as typified by *suicide égoïste* and economic *anomie*, or to the infinite pain of infinite desires inevitably frustrated, as typified by *suicide anémique* [see SUICIDE, article on SOCIAL ASPECTS].

In consequence, Durkheim advocated, as a means of social control over the *anomie* endemic to economic life, the development of professional organizations, in which intense interaction would result in the development of realistic codes of behavior and in their enforcement through constraint upon the conscience of each of the participants.

In *The Rules of Sociological Method*, first published in 1895, Durkheim pointed to another source of deviance: "premature" compliance, like that of Socrates, with currents of the *conscience collective* that have not yet reached the common man. This is essentially a structural approach to deviance; in Merton's famous article (1938), this same approach issued in the concept of deviance as adjustment to the inner contradictions of a given social system rather than as an expression of untamed impulses. Here, however, the resulting concept of social control was not as well elaborated—that is, unless we are satisfied with a more or less explicit concept of social control that limits it to purposive, rational social change aimed at removing these inner contradictions.

In fact, this has long been the approach of American sociology, especially in its early melioristic days. Ross, who is responsible for the popularity of the concept "social control" (it is the title of a book of his, published in 1901), described many of the themes that were to dominate American soci-

ology until the late 1930s. First, there was the concept of the "natural order" derived from the spontaneous meshing of men's personalities with their inherent capacities for sympathy and sociability, on the one hand, and from their sense of justice and capacity for resentment, on the other. Some of these capacities were seen as biological endowments, but the "sense of justice" was thought to be derived from the early development of self not only through interaction as such but also through the resultant internalization of the other, which results in a desire to behave toward others as one would want others to behave unto oneself. Baldwin (1895) was cited by Ross for an early version of a concept that has been attributed solely to Cooley—the "looking-glass self." "Natural order" is thus similar to the state of the primary group as described by Cooley (1909), a state that the Chicago sociologists of the 1920s and 1930s believed was disrupted by urbanization, immigration, breakdown of the small community into competing groups, decline in the efficacy of natural selection (which used to eliminate "moral idiots"), and deviant behavior on the part of men with backgrounds different from those of the original settlers. As examples of natural order Ross cited primitive societies and even such modern settings as the California mining camp, where he thought social and economic equality had reigned before the coming of "adventurers and broken men from all parts of the world" (1901, p. 43).

As Martindale (1966) has noted, social control in Ross's work has many features of social contract, since Ross defined it "as concerned with that domination which is intended and which fulfills a function in the life of the society" (1901, p. 16). Yet when he described social control in action, he fell back upon all the forms of the Durkheimian *conscience collective* that constrain the individual: public opinion, law, belief systems, education, custom, religion, personal ideas (among others, the basic definitions of manhood or womanhood), ceremony, art, personality (i.e., charisma), values, creative elites. The dilemma of the social behaviorist (at least since the time of Lester Ward) is thus clear from Ross's work: on the one hand, there is a Puritan assertion of the importance of the individual, especially of the individual with the proper instincts, as contrasted with the "moral idiots or moral lunatics who can no more put themselves in the place of another than the beast can enter into the anguish of its prey" (*ibid.*, p. 50). These individuals can and should create a moral rational order to replace the natural order rendered un-

viable by greater population density and heterogeneity. On the other hand, Ross maintained that purposive social action, to be effective, must take into account the nonrational elements of social living, which may have Darwinian rationality, if not humanitarian rationality, on their side. But it is never made very clear how we move from the unintended but functional constraint of custom, religion, etc., to an "intended" and even more functional form of social control. Thus, in Ross's discussions of social control we encounter either a description of the elements of social constraint—a field that is equal to the whole of sociology—or pious wishes for rational action, which do not make particularly interesting reading.

W. I. Thomas, as would be expected from one who saw social structure as social definition of the situation, used the concept of social control extensively. He agreed with Ross that once society had moved away from primary group control, the problem was to develop rational controls. Thomas saw social control as emerging from the actions of a new type of man, schooled according to the theories of the philosopher John Dewey and therefore capable of creating new patterns as they were needed (see Thomas & Znaniecki 1918–1920; Thomas 1923). Park and his colleagues (1925) were also concerned over the shift from primary means of social control, such as the family, the neighborhood, and the community, to secondary means of social control, such as the police, the press, the political machine, and the courts. Park pointed out that intended control did not necessarily mean rational control, in terms of the utility of the total community. Drawing on another tradition, Mannheim (1935; 1950) again raised the question of social control through planning and contrasted it with classical liberalism, which could not secure controls over custom and the laws of the market. For him there was a sort of fixed quantum of social control: when some forms of control diminished, it meant that others were increasing, in accordance with what he called the "laws of transmutation." Since controls were unavoidable, the attempt to deny rational, conscious control meant surrendering to nonrational, unconscious controls or to interest-group manipulations.

It is with Parsons (1951) and LaPiere (1954) that we see the concept of social control become systematically limited to the control of deviance. This approach to the concept eliminates from it the ordinary normative component consisting of social structure and, more specifically, the aspects of social structure that are concerned with socializa-

tion. A still more narrow definition is that of *Sociological Abstracts*, which reserves the rubric of social control to penology and the sociology of correction problems.

Parsons' concept of social control is particularly fruitful because it is a systematic attempt to derive consequences from a specific theory of deviance. First, deviance is defined by "its tendency to result either in change in the state of the interactive system, or in re-equilibration by counteracting forces, the latter being the mechanisms of social control. It is presumed . . . that such an equilibrium always implies integration of action with a system of normative patterns which are more or less institutionalized" (1951, p. 250). Leaving aside the problem of error as deviance, Parsons addresses himself to the genesis of deviant motivation, which, in his opinion, is not to be found in "animal spirits" but results from (1) the learning derived from past interaction, (2) specific personality factors, and (3) the pressures and opportunities of immediate interactional situations. When role expectations are frustrated, the motivation of the actor is likely to develop ambivalence. Ambivalence means that the motivation of the actor contains both conforming and alienated components that may be directed at the normative pattern and/or at the relation with the social object (alter). The ambivalence can be oriented to an alter who, as a past satisfactory source of role reciprocity, had become the focus of positive feelings but who now, as a frustrator, becomes a source of negative feelings. Or it can be oriented to the norm that heretofore regulated the relationship between ego and alter and that has been internalized by ego, especially if the norm is seen as responsible for the frustrated role demand.

Ambivalence can result in three courses of action: first, the loss of any attachment to the object or the pattern (a movement from ambivalence to indifference); second, a compulsive expression of only one side of the ambivalence, compulsive because it must inhibit the unexpressed side; and third, the acting out of both the conforming and the alienated sides in contexts that are segregated in time and place. Although Parsons does not say so outright, one is led to the conclusion that the personality system will normally take the third course and try to discharge both components of its motivational states. If this is not possible, the result is more likely to be either compulsive conformity or compulsive alienation. And it is to these two last eventualities that Parsons addresses himself. Passivity-activity, compulsive



conformity versus compulsive alienation, and orientation to patterns versus orientation to social objects are the three major dimensions explored in his analysis of deviant behavior.

This analysis results in the following insights. First, what may on the surface appear to be a conforming disposition may be so shot through with repressed alienation that as a result it is rigid, characterized by attempts to dominate the situation without due regard to the needs of alter. As a result, it is likely to be either deviant or deviance-prone. Second, since illness is a prototype of passive alienated behavior, it must be considered a pattern of deviance. Thus its incidence, form, duration, and so on, become susceptible to sociological analysis. [See ILLNESS; MEDICAL PERSONNEL, *article on PHYSICIANS*.]

Turning to the situational aspects of deviance, Parsons mentions role conflict and value conflict. His insight here is that values, especially of the universalistic-achievement type, lend themselves to deviant orientation because they permit much leeway in interpretation. Thus we have deviant groups able to claim legitimacy as conforming to the "real" values of the society, and this in turn permits deviants to express both sides of their ambivalent motivation: conformity to their own group and its norms, together with alienation from the outside world and its norms. A good example of this is the radical Christian sect.

Although Parsons seems to believe that the grouping together of deviants increases their danger to society, he sees an ultimate source of social control in their ambivalence and in the bridges that these utopian or radical groups maintain to the main value pattern by claiming to be its real defenders. These bridges are the basis for the "selling out" that many members experience after a more or less protracted stay in the deviant organization.

Related to the deviant group's search for legitimacy is the "secondary institution," such as the youth culture or organized gambling, which allows some deviance from the dominant value pattern and yet keeps the participants integrated with this pattern. It is the prototype of institutions that both permit some expressions of alienated feelings and bring the actors back to conformity. [See GAMBLING.] Such, too, is the function of bereavement rituals [see DEATH]. The secondary institution also *insulates* the deviant pattern and limits its impact on the rest of the social system. Insulation as a means of social control is to be contrasted with isolation, the latter aiming to forestall the grouping of deviants. An example of isolation is what

occurs in the treatment of physical illness, for instance, although the treatment of illness as a whole is an example of insulation. And it is precisely in relation to the description of illness as passive deviance that Parsons makes his major contribution to the theory of social control. Taking the psychotherapeutic model as the prototype of the social control response to illness, he sees this process as including the following steps: support, permissiveness, restriction of reciprocation, and, finally, esteem for resuming conformity (Parsons 1951, chapter 7; Parsons & Bales 1955, pp. 38-41).

The relationship of this therapeutic model to the structure of social learning is clear. Social control, when it aims to alter the state of the actor's motivation (rather than merely introduce the situational component of reliable and harsh negative sanctions) will follow the psychotherapeutic paradigm, most often quite unconsciously, and this will happen both at the level of dyadic encounters and at the level of institutional patterns, such as the bereavement rituals mentioned above.

The following analysis of social control stems from Parsons' insights, but it focuses more on the institutional aspects and less on the motivational aspects of the problem. Thus, its main concerns are the institutional patterns that attempt to head off deviance by (1) preventing the buildup of tensions that result in a desire to deviate; (2) reinforcing the desire to conform; (3) making clear what is socially appropriate; (4) discouraging deviation by reliable negative sanctions and rewarding conforming behavior; (5) modifying social patterns to accommodate as much as possible the deviating behavior of actors.

Finally, a description is given of the "fringe organization" as one type of Parsons' secondary institutions, and it is shown that fringe organizations have functional aspects for society even when they seem to give deviants the power of coalition.

### Preventing the buildup of tensions

Prevention of the desire to deviate is, of course, one of the major functions of socialization. What is relevant here is the existence of structural arrangements for draining off tensions that otherwise might trigger deviant behavior. The buildup of tensions in an individual, caused by the inability of role behavior to reduce psychological need dispositions, makes him vulnerable to regression, i.e., to behavior which is not value-oriented and which is incompatible with his normal age, sex, and status roles but which would reduce these tensions. Hence, institutional arrangements have to allow regressive behavior either in isolated situations or

in crowd situations where the behavior gains temporary legitimation through the collective aspects of the behavior, and yet the temporary character of the crowd clearly marks the irrelevance of the behavior for normal situations. This is preferable to the development of latent alienation in conforming behavior.

Primary groups, of course, offer this relative isolation. In fact, role behavior within them requires a capacity to regress and/or to tolerate this capacity to regress in others. Parenthood and conjugal relations are examples of primary group behavior where a capacity to regress—identification with the child, for instance—is an important motivational facility. Family roles offer a most important sanctuary for much tension reduction and behavior that would not be tolerated in a public context. As a matter of fact, one of the most effective agencies of social control is the family, as indicated, for instance, by the finding that marriage is one of the action patterns that has the most positive rehabilitative significance for a young jail or penitentiary releasee (Glaser 1964).

Other modes of releasing tensions are the various forms of entertainment and play through which catharsis of antisocial desires can be secured. The "party" is a sort of institutionalized crowd situation where consumption of alcohol testifies that behavior "under the influence" does not imply a lasting commitment of the self to the pattern being indulged in at that moment. Saturnalia such as the German *Fasching*, carnivals, Halloween, and office parties perform similar functions. It is interesting to note that the wilder parties are typical of the lower-class males, the young of all classes, and the more cosmopolitan of the middle and upper classes. In the lower classes and among the young the problem of role frustration is acute. In the upper classes the opportunities for regression with immunity are at once a mark of class membership and a class asset.

The development of labor discipline in mass industry seems to have paralleled the development of mass entertainment, as if the tensions induced by the former could be endured only if the latter were available to provide release. Urbanism by itself creates many crowd situations where tension release can be secured. The development of the saloon offered an opportunity for adult peer grouping where a certain adolescent irresponsibility could be recaptured, and although it led a certain percentage of workers to alcoholism and family nonsupport, its over-all effects must have been to reduce working-class tensions.

Entertainment, through art, literature, films, or

television, has a social control impact that well illustrates the mechanisms described by Parsons in his psychotherapeutic paradigm. By showing the hero as subject to some foibles, the various art forms offer support and permissiveness. They allow the spectator to have a deviant fantasy without burdening him with isolative guilt. Art must follow canons of taste, and the audience has the duty to take a critical attitude toward the show. Accordingly, regression through entertainment is allowed only because entertainment finally sends the audience back to its cultural responsibility, and through this to its broader social responsibility (for one of the best accounts before Freud of art as social control, see Ross 1901).

Participation in the family, in peer group roles, and in entertainment roles acts as a powerful agency of social control, not only through the surveillance and sanction systems present in the primary group but also through the opportunity afforded the individual to give vent to motivations that cannot be accommodated in secondary group roles. Hence, it is possible to say that all work and no play can make Jack a deviant boy.

Religious services also act in this way, more obviously in the case of revivalist services. Such services not only reward conformity but also, through the function of the church as a "social placer," stress rewards for proper motivation, regardless of actual success. Thus churchgoing decreases the sting of failure by stressing the secondary importance of success in this world. If it did not, failure in this world might impel the individual to give up his conforming motivation. In this context it would be interesting to compare the rates of deviance of the western European working class with those of the American working class. The former has had the benefit of the secular religion represented by revolutionary socialism, while the American working class is largely bereft of this "Catholicism of the poor." Socialism has removed the sting of failure from the urban proletarian condition. It has promised the workers special dignity in the socialist paradise if they keep on being deserving—which, in addition to joining political organizations, means rejecting a difficult, if not impossible, upward mobility and abstaining from the deviance seen as characteristic of that lower-class stratum which Marx called the *Lumpenproletariat*. Thus conformity for the socialist working man is not incompatible with dignity. The American working man, on the other hand, caught in a Puritan ethos of striving and with no sedative for the pains of failure, may well be prone to take the law into his own hands, i.e., strike out against the prevailing



system through various forms of white-collar or blue-collar crime.

A functional and partial equivalent of the Catholicism of the poor is the gambling complex (Bloch 1962, Parsons 1951), which redefines life chances in terms of the "break": thus, one's status in life is not one's own fault, since the same luck that puts you down may raise you up [see GAMBLING].

### Reinforcing the desire to conform

The reinforcement of the motivation to conform was one of Durkheim's major concerns. He saw social control (which he called social constraint) as derived from collective representations; its strength was directly related to the size of the group that shared those representations and to the intensity of the interaction between group members. All social institutions, according to Durkheim, had this function of heightening the strength of collective representations through increasing the rate of interaction. Ritual, on the other hand, increased the sameness of the shared representations at the same time that it grouped together masses of people (Durkheim 1912).

It is not necessary to accept Durkheim's interpretation of the mechanisms whereby this heightening of the constraining power of collective representations takes place in order to accept the statement that ritual does act as a reinforcer of conforming motivation. It makes the actor feel part of an all-enveloping group that shares his attitudes and with which he can think of himself as standing in a primary-type relationship. The actor's good standing in such a group depends upon his continued commitment to the symbolic system evoked by the ritual.

A rationally oriented society is certainly not free of ritual. One of the dimensions of mass entertainment is ritual. Rites of passage, such as commencement exercises, make clear to the individual that he has changed groups and that he must not regress to behavior which might have been proper for the group he has left but is not functional for the groups he has entered.

Ritual operates as an *emblem of membership*. In modern society the commitment to the over-all community, a necessary but not a sufficient condition for commitment to its inner role structure, is stimulated by collective movements that demand the expression of conventional sentiments. These collective movements are triggered by dramatic events such as Lindbergh's flight to Paris, President Kennedy's funeral, or any dramatic event where common feelings are aroused in many people. Press, radio, and television are crucial to this

process. At such times the nation becomes palpable to the individual. How often these movements are necessary and what form they take deserve more research than has been forthcoming heretofore (see, however, Shils & Young 1953; Greenberg & Parker 1965).

Finally, the extension of education has greatly reinforced popular commitment to the nation's values and norms. As modern society allows and demands ever-increasing initiative from the individual rather than binding him to traditional role expectations, the more imperative it becomes for society that individual commitment to values and norms should be thorough.

In situations where role behavior is largely invisible and/or where the alter of the dyadic relation is unable to evaluate whether ego is really upholding his end of the bargain, a long educative process is as necessary for building value controls as it is for imparting technical skills. This is true of most professions. Such occupations as watch and television repair, in which role behavior is also invisible, do not get this expensive value training, with the result that deviance is much more extensive in these trades than in the professions.

### Clarifying the content of conformity

Reinforcement of motivation is of little avail in forestalling deviance if there is ambiguity as to what is conformity and what is deviance. While there are always some areas of anomie in any complex society, an additional hazard is the pressure that the individual generates against the norms through developing idiosyncratic definitions of the situation—definitions that legitimate for him behavior that to others seems clearly deviant (Sykes & Matza 1957).

Primary groups have the capacity, through their close contact with their members and the high rate of communication which prevails among them, to make very clear what is desired or what is proscribed behavior. By confronting the individual's inipient idiosyncratic definitions of the situation with its own definitions, the primary group brings the individual back into line. There are areas of behavior, however, where individuals can become quite isolated from all primary groups, and thus without seeming to be. Marginal groups are especially prone to this invisible isolation. Cressey (1953) has shown how the first-time offender is often a man who becomes socially isolated with respect to a financial problem because the very existence of the problem already involves deviance from status group norms. Around this "isolated deviant problem" as Cressey calls it, the individual devel-





pliance is solved by treating the conformist to the ideal norm as a "hero"—a positive deviant, but a deviant nevertheless. Thus the average member is relieved from the guilt of nonconformity, and there is little development of counteraggression against the ideal norm. Military peer groups may say "Never volunteer," but nevertheless enough soldiers do.

**The law as ideology.** The law has been used, especially in the United States, as a promoter of various kinds of social changes. This is equivalent to treating the law essentially as a statement of ideology. At the present time the federal government of the United States is using law as an instrument for changing the social status of the Negro in the South and, to a lesser degree, in the North. Although Sumner's classic critique (1894) of the "absurd effort to make the world over" still holds true in general, there are nonetheless situations where law judiciously enforced can act as an agent of social change. To take the Negro civil rights issue as an example, it requires, first, the growth of ambivalence in Southern commitment to the caste system and a growing number of supporters for the federal position among the white population. Through judicious enforcement of the law the federal government can create situations in which the Southerners committed to the caste system find enforcing its norms too costly and therefore begin to look on surrender as honorable, because it is into the hands of impersonal but respected powers such as "the court" and "the law of the land." Every failure of the caste system to enforce its norms decreases the power of these norms; every success in enforcing the federal norms strengthens these norms. However, the federal norms, being supported only by a minority, require for their enforcement a reliance upon force that is expensive and partly self-defeating, because force clearly reveals how little support they enjoy from the local mores.

The role of the law as a specifier of the proscribed is fully discussed elsewhere [see *Law*]. One aspect of this role that deserves special mention here is the contradiction in which the legal process finds itself. In the trial the lawyer defends the law against such deviant tendencies of the state as its propensity to abuse its power in order to secure what it believes to be desirable community ends. The lawyer accomplishes this by forcing the state to follow the rules of evidence and the inherent logic of the legal codes. Through his success the majesty and sacred character of the law are reinforced, especially among the middle classes, who

can comprehend the logic of these procedures (though more often through the way in which they are represented by television and the movies than through any actual witnessing of them). On the other hand, it is a natural consequence of the legal process that a certain number of criminal deviants fail to be convicted and punished. Knowledge that this is so weakens the authority of the trial in the eyes of spectator and culprit alike. It increases the pressure for "police justice," or the informal disposition of cases by policemen on the spot. It also increases reliance upon the informal sanctions of the social class system. Acquittal in court comes to mean not innocence, but that the state was unable to prove guilt.

### Reinforcement of the sanction system

Sanctions can be negative or positive, that is, they can punish deviance or reward conformity. Most of the rewarding of conformity takes place through the regular role meshing and role reciprocities of the social structure. It is in this sense that Durkheim is right: all aspects of social structure have a part to play in social control. Here, however, we are primarily concerned with sanctions as deterrents. The effectiveness of a sanction system in discouraging deviant behavior is, we may suppose, directly related to the strength of the sanctions and to their reliability. In fact, if we extrapolate the findings of Lazarsfeld and Thielens (1958) concerning college professors under the cross-pressure of left-wing peer groups and right-wing Congressional investigating committees, we can add that the reliability gradient of sanctions may be more important than their severity gradient. For instance, the difference between the rates of stealing in rural and urban districts, other factors being held constant, can be partly explained by the comparative ease of hiding stolen property in the city because of the lower surveillance capacity of the collectivities in which the urban deviant spends his life. In the city, neither physical proximity nor constant visibility implies a primary group concern for or interest in alter. Thus, detection and negative sanctions are easier to escape in the city even while the symbolic potential of the mere possession of goods increases. There is little knowledge of the owner's roles outside of the property context, and specialized agencies such as the police cannot duplicate the ubiquitous surveillance of long-established neighbors with firsthand knowledge of each other's financial capacities. Hence, the higher the turnover of neighborhood membership, the greater the penetrability of primary groups

(and the greater the ease of departure from them), and the weaker the negative sanctions against all forms of misappropriation of property. For instance, the Chicago racketeer can become a respected real estate operator on the West Coast. To cope with the problem of the invisible past in a mobile society, state and private organizations have developed their own system of investigation, in which the secret personal file has replaced fallible public reputation.

Class discrimination is more effective the closer one comes to the upper-class status group because, at that level, numbers are comparatively small and contacts are maintained through a high rate of sociable interaction and entertainment. As a result, in such circles reputation is more effectively used as a mechanism of social control. For instance, arrest and imprisonment are not a serious disgrace to the lower-class deviant, since he shares in a contraculture, one of the functions of which is precisely to neutralize the impact of such public degradation or loss of self-esteem. But arrest and imprisonment are major catastrophes to a middle-class deviant or to anyone not defended by a delinquent contraculture or subculture.

Thus class sanctions are among the most effective sanctions available for social control. They often act as agents of social control even among those whose self image is one of inured criminality; for instance, it is rare that criminal fathers will not attempt to raise their children to be "straight." Wives and children are the hostages of the class system, and many a criminal has retired from crime, many a conformist has remained a conformist, because of the fear that his family would suffer negative sanctions if his deviance should become known.

The upper class exercises powerful control over the successful entrepreneur of lower-class origins who attempts to continue the self-seeking and deviant orientation that has been so profitable to him in the past. By forcing the upstart's family into a marginal position in which the children are led to judge their parents by standards that the parents cannot learn, the upper class fosters tensions that have literally destroyed many such families. Indeed, the effectiveness of this kind of class system as a sanctioner derives from the fact that it does not debate with the individual as to his guilt or innocence. Instead, it acts upon the basis of reputation and of its members' response to the actions of the individual in question. Since the individual grants legitimacy to the class he aspires to join, his anxiety, generated by the invisibility of his

judges and by the uncertain incidence and magnitude of their sanctions, causes him to undergo anticipatory socialization and to overconform well before he receives the rewards of membership. [See REFERENCE GROUPS.]

Social class sanctions do not require the promise of social mobility to be effective. In fact, a high rate of social mobility may reduce the screening capacity of status groups. In low mobility systems the upper social classes confer esteem on selected lower-class families for their outstanding conformity and thereby raise the latter's standing within their own status groups. A weaker class system, such as *may* exist in the Soviet Union, requires, for the functional equivalent of social class sanctions, the systematic surveillance of the citizen by the state through a large police force, much of it secret, assisted by the citizen militia, factory committees, wall newspapers, and the like. To be as effective as social class sanctions, this kind of surveillance, in its arrest and trial procedures as in all else, must be as one-sided as is the court of social class.

Although social classes share certain broad judgments as to the severity of crimes, they do not agree on the evaluation of certain points of behavior. Each social class has its private culture as well as its private language. For instance, some upper-class groups tolerate extramarital sexual patterns that would be considered deviant by the classes beneath. One of the reasons for this upper-class tolerance is precisely the intolerance that prevails among the classes beneath them. Lower-class deviants are sometimes protected by upper-class persons merely in order to degrade the middle-class standards that these deviants have transgressed. Thus, any exacerbation of the class struggle weakens the efficacy of the class system as an agency of social control.

**Force as negative sanction.** The main asset of primary groups and of the status system as agents of social control is that their sanctions are pervasive in their incidence and impact and that even the deviant is usually highly motivated to maintain group membership. Accordingly, the most drastic sanction at the disposal of these groups is ostracism, which can use force, although it does not usually do so. More often, legal use of force is the monopoly of the state. In any case, force always stands behind the penalties meted out by the apparatus of justice, even when these penalties are merely fines and damages. Force is, of course, the ultimate deterrent and a most powerful means of social control. It can prevent an individual—either



directly, through incarceration or death, or indirectly, through threats—from doing something which the state forbids. In fact, force seems much more effective in prohibiting action than in promoting active conformity. As soon as role tasks take on any sort of complexity, they require a degree of cooperation on the part of the individual that force alone cannot secure. Force can only coerce the visible—the body and its movements—while cooperation, which involves the use of intelligence and good will, is invisible. It is difficult for force to secure cooperation in the sense just described unless the following two conditions are met: (1) members of the deviant's primary groups can be used as hostages, or the deviant can be blackmailed, explicitly or implicitly, through the possibility that he may be expelled from cherished reference groups; (2) the use of force reasserts the legitimate authority of him who uses it by certifying that he takes his responsibilities and convictions seriously.

The deviant can also be compelled by force to participate in roles where force is not a major aspect of the interaction. Such roles may have, as will be seen from the discussion of prisons, a socializing influence. At one extreme is the "brainwashing," the effectiveness of which seems to have been exaggerated, which took place in the Chinese prisoner of war camps in Korea. At the other extreme is the therapeutic relationship which commitment to a mental hospital may permit. But once childhood is over, punitive justice is the only expression of force that most citizens of modernized societies are ever likely to experience. Force certainly has a bad reputation in intellectual circles, and it is true that punitive justice acts at least as much through the stigma carried by its penalties as through its incarceration or rehabilitation programs. By stigmatizing the deviant, punitive justice passes on to him the pervasive sanctions imposed by the primary groups and by the status system. For those who no longer wish to belong to the community of the righteous, these secondary sanctions are meaningless; in fact, to have served a prison term can, in the "negative status group" composed of fellow deviants, be a claim for prestige. But secondary sanctions act as powerful threats for those who conform primarily because they are afraid of downward mobility. The latter not only fear the sanctions of punitive justice but even come to look on avoidance of these sanctions as a reward for conformity. The more gruesome the imagination of what the deviant endures, the stronger the conformist's "relative appreciation" of his own fate. As a matter of fact, the imagination is often worse

than the reality, since those who feel but resist temptation usually stay as far as possible from the areas where they might develop a realistic knowledge of punitive justice.

Those who insist most vehemently that the brand on the deviant should be deep and lasting are likely to be those for whom the tangible rewards of conformity are relatively few and for whom virtue must too often be its own reward. Frequently they are the ones who live closest, in terms of social distance, to the lower class, from which most blue-collar criminals issue. They are middle-class-oriented working-class, the "poor but honest," or lower-middle class freshly arrived from the working class. Thus stigmatizing the criminal reinforces and stabilizes their sometimes tenuous, imperfect, but often hard-earned superiority of status. Members of such lower-middle status groups, with the possible exception of primary and secondary school teachers, have not had enough education to permit them to handle ambiguity with any ease. For them the attempt to sympathize with the deviant does not imply the self-administered status reward implicit in the sophisticated forgiving attitudes that may prevail in status groups further removed from the lower class. More often they see in this attempt a degradation of their efforts to remain conforming and a denial of their social distance from the deviant. Thus, in these groups, punishment of the guilty is a reward for those who have remained innocent.

**The medicalization of deviance.** The diffusion of Freudian thought since the 1920s has had a significant impact upon the distribution of stigma and the incidence of penal sanctions.

First, it has been a force for redefining certain aspects of deviance as illness rather than crime. It is not that the act thus incriminated loses its deviant characteristics, but that the sanction considered suitable to it is no longer of the penal variety. Second, illness is no longer seen as a result of unfortunate or contingent factors, but as related to unconscious motivation. Deviants are no longer so guilty of their crimes, but neither are they so innocent of their illnesses, not only mental illnesses but physical illnesses and accidents as well. And since it is believed that there is an effective therapy for illnesses, whether mental or physical, the deviant who fails to place himself in the hands of the medical profession is guilty at least of enjoying "secondary gains" (Rieff 1959; 1966).

It has been thought by some that the medicalization of deviance has resulted in weakening sanctions to a point where individual responsibility and conformity are endangered. There is no hard evil-

dence on these points. True, the upper classes meet with less tolerance for the vices traditional to their status. On the other hand, judges will often commit the rare middle- and upper-class culprits who come before their courts to a course in therapy rather than to a stay in jail (however, these deviants may well find therapy more painful than jail, even if the social repercussions are less drastic). Lower-class culprits, especially adolescents and young adults, are more likely to be committed to mental hospitals and to find that their sentence is indeterminate. As a result, they may stay in the hospital longer than they would have stayed in jail. Szasz (1963) and other writers have stressed the threat to civil liberties implicit in present commitment and sanity board decisions.

No doubt medicalization has resulted in extending immunity from punishment to certain culprits. However, it would seem that medicalization is one of the most effective means of social control and that it is destined increasingly to become the main mode of *formal* social control. Its advantages are many. For instance, the medical and paramedical professions, especially in the United States, are probably more immune to corruption than are the judicial and parajudicial professions and relatively more immune to political pressure. This is partly due to the fact that the social distance between patient and therapist is greater than that between lawyer and client. Although the doctor has more power in his clinic than the judge has in his courts, the possibility that a patient may be exploited is somewhat minimized by therapeutic ideology, which creates an optimistic bias concerning the patient's fate. American psychiatrists are better trained for their role than American judges are for theirs. Hence, in the competition for administering formal sanctions it is possible that the medical professions will win out. Signs of such a victory would be the transfer of divorce cases to medical and paramedical boards, the end of juvenile courts, and the development of social psychiatry.

The medicalization of deviance denies the deviant the possibility of value legitimacy: If the culprit pleads illness, he thereby certifies his commitment to the dominant values and institutionalized norms and implies that, when restored to health, he will conform. Meanwhile, he must deliver himself into the hands of the medical authorities and follow their instructions in order to regain mental health. Thus the medicalization of deviance results in the political castration of the deviant.

However, medicalization does permit a less wasteful and more flexible handling of deviance.

Except for "organized" crime, such as the provision of gambling opportunities, the drug traffic, and prostitution, it is probable that most "crimes" (that is, forms of deviance classically subjected to court action) are committed by ordinary citizens rather than by members of the underworld. The disadvantage of criminal prosecution is that it can secure reliability of expectations regarding the deviant only at the price of boxing him into a stereotype that he then finds almost impossible to escape. In a sense social order, for its own preservation, *requires* him to become a stereotyped criminal. But forms of deviance that cannot be continually subjected to court action—forms that might be called the criminality of everyday life—must be ignored by the institutions of formal social control and left to the sanction of informal interaction and social class. The medicalization of deviance offers yet another possibility: social pressures on deviance can be increased without boxing the deviant into as rigid a category as "criminal."

The strength of informal sanctions is declining because of the increase in geographical mobility and the decrease in the strength of the traditional status groups. Medicalization offers a substitute method of controlling deviance—a method that is being used increasingly by universities, large corporations, professional organizations, social work agencies, and even by trade unions to control their own members and clients. The increasing commitment of the general population to absolute values—a commitment due partly to increasing education—makes a therapeutic approach more practicable. But the growing complexity of the division of labor and the increase in the zones of autonomy that are possible in the new bureaucracies (Crozier 1963) make it more imperative that the decisions which are taken by employees should be "reasonable." It is the problem of the pilot with the atomic bomb. An even greater sector of the population works with similar responsibilities. Thus social control becomes more humane and forgiving, but perhaps also more relentless and pervasive.

Although mental illness carries a stigma that may create resistance in the deviant to the role in which medicalization has cast him, it may also compel the referring authority to give the "new" patient a second chance, if the medical profession certifies him improved or cured. This can decrease the deviant's fear of being stereotyped and so reduce his resistance to treatment.

Because the medicalization of deviance is imperfectly institutionalized, there are still problems in the handling of medical files, and past treatment may become an unnecessary bar to certain



positions. These problems are likely to be worked out as institutionalization proceeds and as scientific knowledge of personality potential and effective role requirements becomes more widespread.

The medicalization of deviance has so far provided greater flexibility of response to deviance, together with more effective detection. This is partly because those who support and implement the medicalization of deviance now oppose many of the existing institutionalized arrangements for social control. Indeed, problems of control have been created in institutions such as prisons, communities for juvenile delinquents, and mental hospitals by the frequent identification of the professional with the deviant and their joint hostility to the petty bourgeois-oriented custodial staff. The unconditional rewards the professional staff offers to the deviants in order to establish therapeutic dyads, and also in order to degrade implicitly the status of the custodians, demoralize the latter, who see deviants secure favors and commercium with upper status groups that are denied to them. Their response is to sabotage the therapeutic program and try to take back the favors.

### Modification of social patterns

Social control is not exclusively concerned with trying to change deviant motives; it also operates by systematically isolating the deviant from the rest of society through the agency of such specialized organizations as hospitals, prisons, and police forces, or through that of certain relatively informal modes of organization capable of monopolizing much of the deviant's behavior and of limiting, as well as patterning, his relations with the outside world. Examples of such organizations are the doctor-patient dyad, the underworld, the delinquent gang, bohemian coterie of various kinds, and political and religious sects of such an extremist tinge that they are outside the mainstream of political and religious thought.

Hospitals and prisons, whatever else they may be, are certainly not models of conformity. Rather, they resemble such classic forms of institutional deviance as gambling, prostitution, and drug addiction, which operate under the informal supervision of the police rather than in spite of them. A mental hospital, from this point of view, is a place where mental illness is *allowed*. Another major aspect of formal organizations specializing in social control is that they control the staff as well as the inmates. Frequently, as in mental hospitals and penitentiaries, the staff lives on or near the grounds and its contacts with the outside world are few. [See SOCIAL CONTROL, article on ORGANIZATIONAL ASPECTS.]

Most work with deviants implies a capacity to identify with them that is above the average level of human empathy, and certainly a capacity to overcome the average distaste aroused by the stereotype of the deviant. Often this capacity is based upon strong deviant tendencies within the staff member, as described by the folkloric statement concerning the police: "It takes one to find one." Similar remarks could be made regarding mental hospital staff (especially professional, but also non-professional), but no solid evidence has yet been collected in this area. However, since all occupational roles that succeed in securing a high level of personality commitment must operate to reduce the deviant potential of their incumbents, the statement that any agency of social control also controls its staff implies that the staff brings to the agency a greater propensity to deviance than is met with in the average breadwinner. Impressionistic evidence of this seems strong in the cases of vice squad detectives, psychiatrists, and personnel of organizations devoted to the rehabilitation of alcoholics.

Research on this problem is urgently needed, but such research would create very serious problems of rapport and public relations for the sociologist. Nevertheless, it is useful, in order to understand the structure of these agencies, to think of them as organizations where dilemmas of administration and policy are resolved in terms of the staff's needs rather than the inmates', even though this is not intended. The reasons for this are threefold. First, roles specializing in social control deal by definition with people who are frequently recalcitrant to, and are not deemed competent to evaluate the worth and necessity of, the restraining and/or therapeutic efforts directed at them. Second, there is a tendency for the stigma that applies to the deviant to be extended to the organizations specializing in control of the deviant. As a result of both these circumstances, formal social control operates at a very low level of visibility. The receivers of control are deemed incompetent and inferior to their controllers. The rest of society is not very interested in what is done to them as long as physical isolation is successfully maintained, and the standards of middle-class pity are applied. Third, although physical isolation is comparatively easy to achieve because it depends upon the known capacities of the human body (and also because a good percentage of inmates have no real intention of leaving), both rehabilitation and therapy remain problematic because so little is known about either. In prisons, especially, the inmate peer group has closer contact and control over the individual inmates than the staff either can develop or wishes

to develop, and these peer groups will resist the attempts of the staff to reform the inmates. Thus the relation between the staff and the inmates is reminiscent of the relations between the police and organized crime. The staff's administrative and rehabilitative policies will follow a combination of ideological and self-gratificatory lines. Colleges and universities are not very different in this respect.

**The prison as socializer.** Under the conditions described above, it is surprising that the recidivism rate of released prisoners oscillates between 30 and 40 per cent (Glaser 1964). True, some of this is accounted for by the ex-convict's better techniques for evading arrest, just as many of the patients who never return to the mental hospital have learned to be sick at home. And there is a certain percentage of prisoners (especially wife murderers and first-time embezzlers) who reject convict values and never intend to repeat their offenses (in American prison slang these prisoners are known as "square Johns"). Wheeler (1961) has shown that there seems to be a U-curve of conformity to social values on the part of the inmates who do not belong either to the class of "square Johns" or to that of "gorillas" (seasoned hoodlums). As the inmate nears release, he seems to make some attempt to reorient himself to the values of the outside world. As one would expect from Parsons' analysis of the ambivalence of deviant attitudes, a homogeneous and definitive deviant motivational structure is as rare as a homogeneous and definitive conforming motivational structure. Hence the desire and capacity of the convict to rejoin those who conform are greater than had been foreseen by the critics of the prison system.

The major obstacle to rehabilitation seems to lie in the transition from prison to the outside world. The difficulties center in two areas: (1) the fantasies about civilian life that are elaborated in the prison concerning jobs, income, and women insure that the ex-convict will be disappointed with conforming roles, especially since he can rarely secure one of the more attractive jobs immediately upon release; (2) the loss of primary group support that is involved in quitting the underworld to join the conforming community leaves the ex-convict without support while he is trying to break through the public stereotypes—although there is impressionistic evidence that the rejection he anticipates may be more of a problem than the rejection he experiences. The new trend in rehabilitation seeks to cope with these problems by creating "halfway houses" that facilitate and guide the ex-convict's return to society and offer him the support of a peer group composed of people who are in the same situation and who are trying to succeed in it. This

form of peer group support is lacking in traditional parole procedures. [See *PENOLGY*.]

The other way in which prison life may alter the motivation of the inmate, especially the young inmate, is to develop in him the personality attributes that are functional to the successful discharge of conforming roles. The ideals of the "right guy" (Sykes 1958) are those of self-control, self-reliance, avoidance of self-pity, not getting into trouble, and avoiding or minimizing involvement with homosexuality. For the young adults who make up the bulk of prison commitments, the capacity to internalize these attitudes would imply a capacity to resist delinquent peer group pressures and the impulses to aggressive acting out that often caused their incarceration in the first place. It is true that the result of modeling oneself after the "right guy" might be a deeper commitment to underworld values and—perhaps more important—to the underworld as a collectivity and to its role structure. But some of this commitment to underworld values might be, if one accepts Wheeler's U-curve theory, more conducive to adjustment than effective internalization would be. This is because commitment to the ideology of the convict group in the prison is more a flag of membership and a boundary maintainer against the authority structure than a real commitment to action on lines supported by this ideology. Accordingly, "prisonization" (that is, getting adjusted to the conditions of prison life) is not the equivalent of criminalization (Glaser 1964). For those who internalize both the underworld value system and the behavioral skills symbolized by the concept of the "right guy," the changes in personality structure may well result in a better adjustment to the outside world through a greater conformity to its basic rules of behavior.

**The fringe organization.** An organization that promotes policies sharply at variance with the political, religious, or moral principles of the society, or that imply at least a drastic change in the hierarchy of social values, may properly be called a fringe organization. The concept is a development of Parsons' "secondary institution."

The fringe organization claims a special relation to the dominant values even though, in fact, it deviates from the established norms and role expectations. It tends to attract deviants whose field of deviance may be quite remote from the overt goals of the organization. What the deviant finds in his membership is a re-evaluation of his personality. He can now reject the outside rejecters. His superego pains are lifted and transformed into anxiety to do well in his new group. [See *SOCIAL MOVEMENTS*.]

As the member achieves good standing in the



fringe organization, he is granted esteem, and this consolidates his learning. However, the support and esteem that legitimate the member's ego in his own eyes also increase the legitimacy for him of those needs of his which are not integrated with the fringe organization's role structure. Soon the member learns dissent. He discovers the discrepancy between the group's utopian values and its organizational limits. He comes to realize that the out-groupers are neither better nor worse than the in-groupers. His growing participation in activities involving out-group members (for the adolescent and young adult these activities are often love affairs) increases his resistance to the fringe organization's attempt to monopolize the whole of his life. Soon he is ready to leave the organization and rejoin the wider community. His experience in the fringe organization has served as a therapeutic experience that permitted him to work out pre-Oedipal, Oedipal, or post-Oedipal problems that formerly made it difficult for him to assume his regular age, sex, and status roles.

The fringe organization thus serves society as a resocialization agency. But its services must be publicly unrecognized if it is to remain effective. Harassment by the conforming world is an asset to the fringe organization, which therefore engages in activities capable of triggering sufficient harassment to warrant feelings of martyrdom and hostility but not sufficient to end in the destruction of the group. Hence, the group must plan its deviance as a combined effort. This makes its protest more effective but at the same time compels it to follow the norms of intergroup conflict that are proper to the society at large. Thus the fringe organization becomes in some ways more dangerous to society and in other ways more reasonable and more predictable than would be the many unpredictable role failures of its members if the latter were let loose upon the world as random individuals.

Behind the screen that the police and public interpose between the fringe organization and the rest of society lies a deviant social system that might even represent an institutionalization of a higher level of value commitment, and so prove itself superior to the existing social system. For a society can be wrong in terms of its objective possibilities; the deviant may be right, and the better often begins as the enemy of the good. More often than not, however, the fringe organization will be merely a group that enrolls the emotionally handicapped under the flag of utopia. Yet it may also be a group that, like the first Christian sect, may teach the world a way to live in society with less bloodshed, less violence, less contempt. By tolerating the

existence of the fringe organization, society leaves the way open for its own renewal.

Social control, as a specific topic for sociology, seems to be less popular now than in the past, when it was closely related to a concept of the source of deviance as being essentially located in the individual. But as the source of deviance has come to be located in the contradictions of the social system, the idea of controlling deviant individuals has become somewhat repugnant, and social reform itself has begun to look more than a little naive to a sociology penetrated with the knowledge that deliberate change always has unanticipated consequences.

Social control was once a concept which covered many areas that later became independent specialties—areas like penology, medical sociology, and the sociology of law. Even such large, amorphous areas as social change can be said to overlap with the area of social control. Hence there is now less interest in a topic which, if it is to be treated as something other than a summary of the various specialties, requires a level of theoretical generality that may seem beyond our present reach. However, it is possible that the concept has been neglected for two reasons which are rooted in the structural conditions of sociology as an intellectual enterprise. First, there is the fact that the study of social control would seem to require a combination of professional skills both in the sphere of individual psychology and in social system analysis. At any rate, such a combination will be necessary if we are to secure (1) a better understanding of the meaning of role participation for the economy of the personality (for instance, in the tension-inducing and tension-reducing role functions that may have a direct effect in promoting or decreasing deviant propensities); and (2) a better understanding of adult socialization patterns, so that we can ascertain the impact of social control agencies in fostering motivational changes that go beyond mere increased awareness of expediency [*see* SOCIALIZATION, *article on* ADULT SOCIALIZATION]. This would require, among other things, research into the life patterns not only of members of total and of fringe organizations but also of those who have left them, regardless of the costs of tracing them and of getting *reliable* indications of their conforming and nonconforming behavior.

The second structural handicap to the development of a theory of social control is the fact that it is an area which is rent with ideological debate. In particular, there is a struggle for the leadership of social control agencies between therapy-oriented

professionals and supporters of the classical judicial apparatus. Sociologists have sentimental and economic ties with the first group, and this makes it harder for them to maintain scientific detachment. Some, who feel more acutely the possibilities of conflict between their status group allegiance and their allegiance to objectivity, try to resolve the problem by flight from the field. Others will no doubt struggle on and let the facts speak for themselves.

JESSE R. PITTS

[Directly related are the entries CRIME; CRIMINAL LAW; DEVIANT BEHAVIOR; INTERNMENT AND CUSTODY; LAW; NORMS; PENOLOGY; PUNISHMENT; SANCTIONS; SOCIETY; VALUES. Other relevant material may be found in CAPITAL PUNISHMENT; DELINQUENCY; DRUGS, article on DRUG ADDICTION: SOCIAL ASPECTS; GAMBLING; INCEST; LEGAL SYSTEMS; POLICE; PROSTITUTION; PSYCHIATRY, article on the RELIGIO-PSYCHIATRIC MOVEMENT; RITUAL; SECTS AND CULTS; SOCIAL MOVEMENTS; UTILITARIANISM; and in the biographies of BALDWIN; BENTHAM; COOLEY; DURKHEIM; FREUD; MALINOWSKI; MANNHEIM; PARK; ROSS; SUMNER; SUTHERLAND; THOMAS.]

#### BIBLIOGRAPHY

- BALDWIN, JAMES M. (1895) 1906 *Mental Development in the Child and the Race: Methods and Processes*. 3d ed., rev. New York and London: Macmillan.
- BARBER, BERNARD 1963 Is American Business Becoming Professionalized? Analysis of a Social Ideology. Pages 121-145 in Edward A. Tiryakian (editor), *Sociological Theory, Values, and Sociocultural Change: Essays in Honor of Pitirim A. Sorokin*. New York: Free Press.
- BLOCH, HERBERT A. 1962 The Gambling Business: An American Paradox. *Crime and Delinquency* 8:355-364.
- CAMERON, MARY OWEN 1964 *The Booster and the Snitch: Department Store Shoplifting*. New York: Free Press.
- COHEN, ALBERT K. 1966 *Deviance and Control*. Englewood Cliffs, N.J.: Prentice-Hall.
- COOLEY, CHARLES H. (1909) 1956 *Social Organization: A Study of the Larger Mind*. In Charles H. Cooley, *Two Major Works: Social Organization and Human Nature and the Social Order*. Glencoe, Ill.: Free Press. → Each title reprinted with individual title page and pagination. Separate paperback editions were published in 1962 by Schocken.
- CRESSEY, DONALD R. 1953 *Other People's Money: A Study in the Social Psychology of Embezzlement*. Glencoe, Ill.: Free Press.
- CROZIER, MICHEL (1963) 1964 *The Bureaucratic Phenomenon*. Univ. of Chicago Press. → First published as *Le phénomène bureaucratique*.
- DURKHEIM, ÉMILE (1895) 1958 *The Rules of Sociological Method*. 8th ed. Edited by George E. G. Catlin. Glencoe, Ill.: Free Press. → First published as *Les règles de la méthode sociologique*.
- DURKHEIM, ÉMILE (1912) 1954 *The Elementary Forms of the Religious Life*. London: Allen & Unwin; New York: Macmillan. → First published as *Les formes*
- élémentaires de la vie religieuse, le système totémique en Australie*. A paperback edition was published in 1961 by Collier.
- GLASER, DANIEL 1964 *The Effectiveness of a Prison and Parole System*. Indianapolis, Ind.: Bobbs-Merrill.
- GREENBERG, BRADLEY S.; and PARKER, E. B. (editors) 1965 *The Kennedy Assassination and the American Public: Social Communication in Crisis*. Stanford Univ. Press.
- LAPIERE, RICHARD T. 1954 *A Theory of Social Control*. New York: McGraw-Hill.
- LAZARSFELD, PAUL F.; and THIELENS, WAGNER JR. 1958 *The Academic Mind: Social Scientists in a Time of Crisis*. A report of the Bureau of Applied Social Research, Columbia University. Glencoe, Ill.: Free Press.
- MALINOWSKI, BRONISLAW (1929) 1962 *The Sexual Life of Savages in North-western Melanesia: An Ethnographic Account of Courtship, Marriage, and Family Life Among the Natives of the Trobriand Islands, British New Guinea*. New York: Harcourt.
- MANNHEIM, KARL (1935) 1940 *Man and Society in an Age of Reconstruction: Studies in Modern Social Structure*. New York: Harcourt. → First published as *Mensch und Gesellschaft im Zeitalter des Umbaus*.
- MANNHEIM, KARL 1950 *Freedom, Power, and Democratic Planning*. New York: Oxford Univ. Press. → Published posthumously.
- MARTINDALE, DON A. 1966 *Institutions, Organizations, and Mass Society*. Boston: Houghton Mifflin.
- MATZA, DAVID 1964 *Delinquency and Drift*. New York: Wiley.
- MERTON, ROBERT K. (1938) 1957 *Social Structure and Anomie*. Pages 131-160 in Robert K. Merton, *Social Theory and Social Structure*. Rev. & enl. ed. Glencoe, Ill.: Free Press. → First published in the *American Sociological Review*.
- PARK, ROBERT E.; BURGESS, ERNEST W.; and MCKENZIE, RODERICK D. 1925 *The City*. Univ. of Chicago Press.
- PARSONS, TALCOTT (1948) 1955 *Illness and the Role of the Physician: A Sociological Perspective*. Pages 609-617 in Clyde Kluchohn and Henry A. Murray (editors), *Personality in Nature, Society, and Culture*. 2d ed., rev. New York: Knopf.
- PARSONS, TALCOTT 1951 *The Social System*. Glencoe, Ill.: Free Press.
- PARSONS, TALCOTT; and BALES, ROBERT F. 1955 *Family, Socialization and Interaction Process*. Glencoe, Ill.: Free Press.
- PITTS, JESSE R. 1961 Introduction [to Part Three]. Pages 685-716 in Talcott Parsons et al. (editors), *Theories of Society: Foundations of Modern Sociological Theory*. New York: Free Press. → See especially pages 701-716.
- RIEFF, PHILIP 1959 *Freud: The Mind of the Moralist*. New York: Viking.
- RIEFF, PHILIP 1966 *The Triumph of the Therapeutic: Uses of Faith After Freud*. New York: Harper.
- ROSS, EDWARD A. (1901) 1928 *Social Control: A Survey of the Foundations of Order*. New York and London: Macmillan.
- SHILS, EDWARD; and YOUNG, MICHAEL 1953 *The Meaning of the Coronation*. *Sociological Review* 1:63-81.
- SUMNER, WILLIAM GRAHAM (1894) 1911 *The Absurd Effort to Make the World Over*. Pages 195-210 in William Graham Sumner, *War and Other Essays*. New Haven: Yale Univ. Press.



- SYKES, GRESHAM M. 1958 *The Society of Captives: A Study of a Maximum Security Prison*. Princeton Univ. Press.
- SYKES, GRESHAM M.; and MATZA, DAVID 1957 *Techniques of Neutralization: A Theory of Delinquency*. *American Sociological Review* 22:664-670.
- SZASZ, THOMAS S. 1963 *Law, Liberty and Psychiatry: An Inquiry Into the Social Uses of Mental Health Practices*. New York: Macmillan.
- THOMAS, W. I. (1923) 1964 *The Unadjusted Girl: With Cases and Standpoint for Behavior Analysis*. Boston: Little.
- THOMAS, W. I.; and ZNANIECKI, FLORIAN (1918-1920) 1958 *The Polish Peasant in Europe and America*. 2d ed. 2 vols. New York: Dover.
- WHEELER, STANTON 1961 Socialization in Correctional Communities. *American Sociological Review* 26:697-712.

## II

## ORGANIZATIONAL ASPECTS

All social units control their members, but the problem of control in organizations is especially acute. Organizations are social units which serve specific purposes. They review their own performances and restructure themselves. In this way they differ from natural social units such as the family, ethnic groups, or the community. The artificial structure of organizations and their concern with performance, as well as their tendency to be larger than natural units, make informal control insufficient and primary identification inadequate. Hence, organizations tend to require formally institutionalized allocation of rewards and penalties to enhance compliance with their norms, regulations, and orders. As a rule, organizations cannot rely on most of their participants to carry out assignments voluntarily. The participants need to be controlled.

For organizational means to fulfill their control function they have to be allocated differentially, so that performances desired by organizational norms will be rewarded while undesired performances will be punished. The allocation of means by the organization independently of performances, such as rewarding people for being white rather than Negro or male rather than female, does not directly enhance organizational control. Allocation according to such irrelevant criteria is more common in the less developed countries and is one of the reasons for the low effectiveness of organizational control in these countries.

The means of control used by an organization can be exhaustively classified into three analytic categories: physical, material, and symbolic. The use of a whip or a lock is physical in the sense that it affects the body; the threat to use physical means is viewed as physical because the effect on the subject is similar to that of the actual use. The

application of physical means for control purposes is here referred to as *coercive power*. Material rewards consist of goods and services; the granting of symbols (for instance, money) that allow one to acquire goods and services is classified as material because the effect on the recipient is similar to that of material means. The use of material means for control purposes is here referred to as *utilitarian power*. Symbols the use of which does not constitute a physical threat or a claim on material rewards should be viewed as pure symbols. These include normative symbols such as those of prestige and esteem and normative social symbols such as those of love and acceptance. (When physical contact is used to symbolize love or material objects to symbolize prestige, such symbols are viewed as normative because their effect on the recipient is similar to that of pure symbols.) The use of symbols for control purposes is here referred to as *normative* or *normative-social power*. Normative power is exercised by those in higher ranks to control the lower ranks directly; normative-social power is more commonly used indirectly—for instance, when the higher in rank use an individual's peer group to control him, as a teacher might do in a classroom.

The use of various classes of means for control purposes—power, in short—has different consequences in terms of the nature of the discipline elicited. All other things being equal, the use of coercive power, in most cultures at least, is more alienating to those subject to it than is the use of utilitarian power, and the use of utilitarian power is more alienating than the use of normative power. Or, to put it the other way around, normative power tends to generate more commitment than utilitarian power, and utilitarian power more commitment than coercive power. The application of symbolic means of control tends to persuade people, that of material means tends to build up their self-oriented interest in conforming, and that of physical means forces them to comply.

Most organizations most of the time use more than one kind of power. The kinds of power used vary according to the ranks of the participants to be controlled. Organizations tend to apply the less alienating means of control to their higher participants. Coercive power is as a rule applied only to lower participants: inmates are locked up if they try to escape. Higher participants are more often rewarded materially in order to increase their performances. It is therefore essential, for sociological purposes, to compare participants of the same rank in different kinds of organizations or of different ranks within the same organization. Otherwise one

- SHILS, EDWARD A. 1948 *The Present State of American Sociology*. Glencoe, Ill.: Free Press.
- SHILS, EDWARD A. 1961 Centre and Periphery. Pages 117-130 in *The Logic of Personal Knowledge: Essays Presented to Michael Polanyi*. London: Routledge; New York: Free Press.
- SOROKIN, PITIRIM A. 1928 *Contemporary Sociological Theories*. New York: Harper. → A paperback edition was published in 1964 as *Contemporary Sociological Theories Through the First Quarter of the Twentieth Century*.
- SPIRO, MELFORD E. 1961 An Overview and a Suggested Reorientation. Pages 459-492 in Francis L. K. Hsu (editor), *Psychological Anthropology: Approaches to Culture and Personality*. Homewood, Ill.: Dorsey.
- STINCHCOMBE, ARTHUR L. 1959 Bureaucratic and Craft Administration of Production: A Comparative Study. *Administrative Science Quarterly* 4:168-187.
- UDY, STANLEY H. 1959 "Bureaucracy" and "Rationality" in Weber's Organization Theory: An Empirical Study. *American Sociological Review* 24:791-795.
- WEBER, MAX (1904-1905) 1930 *The Protestant Ethic and the Spirit of Capitalism*. Translated by Talcott Parsons, with a foreword by R. H. Tawney. London: Allen & Unwin; New York: Scribner.
- WEBER, MAX 1920-1921 *Gesammelte Aufsätze zur Religionssoziologie*. 3 vols. Tübingen (Germany): Mohr. → Volume 1: *Die protestantische Ethik und der Geist der Kapitalismus. Die protestantischen Sekten und der Geist des Kapitalismus. Die Wirtschaftsethik der Weltreligion*. Volume 2: *Hinduismus und Buddhismus*. Volume 3: *Das antike Judentum*. Volume 2 was published in 1958 by Free Press as *The Religion of India: The Sociology of Hinduism and Buddhism*.
- WEBER, MAX (1922) 1954 *Max Weber on Law in Economy and Society*. Edited, with an introduction and annotations, by Max Rheinstein. Cambridge, Mass.: Harvard Univ. Press. → First published as Chapter 7 of Weber's *Wirtschaft und Gesellschaft*.
- WHITING, JOHN W. M. 1954 The Cross-cultural Method. Volume 1, pages 523-531 in Gardner Lindzey (editor), *Handbook of Social Psychology*. Cambridge, Mass.: Addison-Wesley.

## SOCIAL INTEGRATION

See under INTEGRATION.

## SOCIAL MOBILITY

Social mobility is the movement of individuals, families, and groups from one social position to another. The *theory of social mobility* attempts to account for the frequencies with which these movements occur.

The study of social mobility relates a present to a past social position. It thus forms part of the more general study of *social selection*, i.e., of how people get distributed into different social positions. It is, however, hardly possible to study effectively the influence of past social position except in the context of other influences that determine the in-

dividual's present social status. In practice, it has become increasingly difficult to separate the fields of social mobility and social selection. No rigorous separation has been attempted in the following discussion.

Studies of *intergenerational mobility* compare the social positions of parent and offspring; studies of *career mobility* compare the social positions of the same individual at different times. *Group mobility* is concerned with changes in the social position of groups possessing a relatively homogeneous status (for example, castes, intellectuals, artisans).

In the context of mobility studies, *social position*, or *social status*, signifies a certain rank with respect to the possession of goods (values) esteemed and desired by most members of a society. The changes in social position that interest the theory of social mobility are primarily variations in occupation, prestige, income, wealth, power, and social class. A high or low rank in one of these values is often associated with a roughly corresponding rank in most of the other values; consequently, position with respect to one of these values, and more especially a constellation of them, provides a measure of what in many societies is viewed as success in life. Studies in social mobility do not usually concern themselves with the possession of aesthetic, moral, and spiritual values. This is presumably due to the supposition, correct for most societies, that these goods do not measure "success in life." Nor does their possession seem to lead, except in a limited number of societies, to the attainment of those material goods whose pursuit is more evident in human behavior and whose possession tends to limit the amount possessed by others and to provide opportunities for—or at least the illusion of—control over one's own and others' destinies.

### Mobility and moral critiques of society

Political and moral critiques of society have both inspired and been inspired by studies in social and, more particularly, intergenerational mobility. This is understandable in view of the concern of intergenerational studies to show the relation of an individual's life chances (income, occupation, prestige, etc.) to the social circumstances (parental status) in which he is brought up. Casual observation, historical studies, and quantitative inquiry have long made it evident that despite considerable intergenerational movement up and down the social ladder, many children of high-status and low-status parents retain, when adults, approximately the same status levels as their parents. Egalitarian sentiments are affronted by a social process that appears to condemn many children to the inferior life cir-



cumstances of their families or to guarantee to other children, by virtue of the more favored position of their families, a high degree of fortune. Observation of these continuities in family social position has led to the characterization of societies as "open" or "closed" according to the degree to which the adult status of offspring is independent of (*open society*), or dependent on (*closed society*), the social status of the parents.

Independence is sometimes interpreted, in this context, in its strictly statistical sense: societies are egalitarian (open, democratic) to the extent that children coming from very different family backgrounds have the same probability of achieving a specified status level. This criterion, however, could be met without any commonly accepted view of social justice being fulfilled. This distribution of social positions might equally well be achieved were children assigned to occupations or income levels by a state lottery or by a purely whimsical procedure of an absolute ruler.

Societies are also sometimes viewed as being more or less open according to the frequency with which social mobility occurs. This criterion is not consistent with the criterion of statistical independence. The maximum frequency of social mobility occurs (for example, in a two-status system) when all the children of high-status parents fall to low status and all children of low-status parents achieve high status. This entails certain consequences of dubious accord with commonly accepted notions of social equity: offspring of low-status parents must resign themselves to seeing *their* children excluded from high status; and the future social position of children is as fully fixed by that of their parents as in the most rigid caste society.

Evidently, moral and political critiques of society must rest not on distributions of children's status by parents' status (see Tables 1-4) but on the nature of the processes that produce these distributions. Late eighteenth-century and nineteenth-century writers on these problems were concerned in their critiques of society with the institutional and personal factors involved in mobility; they aspired toward a society in which merit and talent were rewarded and opportunities for their development and exercise were freely available. Contemporary writers have by no means abandoned this theme; but with the increased availability of quantitative data they have been more disposed to substitute summary statistical indices of the frequency of mobility for critical analysis of the total process of social selection. This seems to assume either that certain frequencies are in themselves desirable without regard to the institutional and personal

factors involved in social selection or that the frequency of mobility in itself provides secure knowledge concerning the nature and moral legitimacy of the selection system. Neither of these assumptions appears justified.

### The mobility table

A central feature of most quantitative studies of intergenerational mobility is the mobility table (see Tables 1-4). An understanding of the technical problems associated with it is essential to the interpretation of the large body of data such tables have made available.

**The mobility scale.** If no additional assumptions are made, the social positions (e.g., occupations) between which movement may occur remain a series of discrete, qualitatively different categories. The study of social mobility is, however, closely related to questions concerning social achievement and, more particularly, concerning the role of the parental social position in determining filial status. Interest attaches, then, to the amount of *upward* and *downward* mobility (or conversely, the amount of nonmovement, or *inheritance*). The terms "upward" and "downward" imply an *ordering* of the categories along some quantitatively defined axis. When the social positions are defined in terms of (occupational) prestige or income level, an ordering is more readily made, although in the first case not without difficulty. Often, however, the social position categories represent occupational classes (professional, entrepreneurial, clerical, skilled, etc.) or social classes (upper, middle, lower) arrived at, not by one clear-cut criterion (e.g., income, education), but by a mixture of criteria that often leaves the meaning of ordering on a *single* axis, and hence the interpretation of the data, in doubt. The difficulty is not entirely resolved by reducing the multiple criteria to a single score, since a considerable measure of arbitrariness enters into the weight assigned to each criterion in the total score.

Even if a satisfactory ordering is achieved, the categories may not mean—and in the empirical literature generally have not meant—that the social positions represent *scale* positions. It is, therefore, not always possible to say whether movement ("distance") from social position A to social position B is as great as, greater than, or less than, movement from B to C. Recourse is sometimes had to speaking of the number of "steps" (ordered classes) through which movement has occurred. Clearly, however, these steps do not necessarily have the same significance, and this creates difficulties in relating the probability of movement be-

tween two positions to the "distance" separating them.

**Mobility and number of social categories.** A fundamental datum of the mobility table, the proportion of sons who manifest upward or downward mobility (or conversely, inheritance of the parental position), can be made larger or smaller by a simple change in the number of occupational or other classes representing the range of social positions. Thus, the amount of mobility (or inheritance) that a mobility table reveals is dependent on how fine (or broad) the social position categories are—a point that requires particular attention when inter-society comparisons are being made. The tendency toward arbitrariness in the number of occupational or other classes employed is increased both by the practical necessities of research (data availability, manageability) and by the use of multiple criteria in constructing the social-position classes.

**Mobility within the life-span.** Social position often varies during the life-span of the individual. It would, therefore, be desirable to make comparisons of the social position of father and son at several time points during their careers. This would also throw light on career mobility and bring its investigation into closer relation to intergenerational studies. Practical difficulties, especially in the specification of the father's status at several time points, have generally discouraged the use of this procedure. If only a single time point in each career is used to fix the social position of the father and his son, a choice must be made appropriate to the objectives of the study.

When the aim is simply to measure the frequency with which sons attain a higher, lower, or similar status to that of their fathers, it is reasonable to match their occupations at corresponding ages, preferably a fairly mature age, when the occupational career has become relatively stable. Other choices are to compare the highest status levels achieved in each generation and to compare the occupations pursued over the longest number of years.

On the other hand, when analysis is directed toward understanding why the son attains the particular status that distinguishes his position, quite different considerations enter. The question is then which point in the parental occupational career best aids in the prediction of the filial status. Consequently, the investigator must choose a hypothesis to be tested; the question of comparability in the ages of father and son is not involved.

**Sampling problems of father-son studies.** Intergenerational mobility tables are usually constructed by obtaining from a sample of subjects (sons)

their occupations and those of their fathers. But the probability of major interest is the probability that a son will attain a certain status given that the parent has a certain status. From this standpoint, and some others, it would be preferable to select a sample of fathers and trace the occupations of *all* of their sons, since the desired probability (when calculated from a table based on a sample of sons) is subject to error. This procedure is generally avoided because of the considerable age the fathers would have had to attain in order to ensure that the sons have likewise attained a fairly mature age and stable occupation. In addition, the fathers who have survived to that age would be a biased sample of the fathers of the current generation.

**Adjusting for size of occupation.** The mobility table provides an estimate of the relative frequency with which the sons of any particular class of fathers will be found in their fathers' occupation (inheritance) or in an occupation of higher or lower status (mobility). However, the entries in the mobility table reflect not only the effect of the father's status on the son's occupational locus but also the effect of the size of each occupational group. The probability that the son of a cabinet-maker will also become a cabinetmaker is a function not only of the special advantages and motivations that may accrue to him from the nature of the paternal occupation (and family circumstances correlated with it) but also of the number of cabinetmakers that the society requires or supports. It is often desirable, then, to separate the component of mobility that is due to the current occupational distribution from the component that reflects the influence of the parental status. One way of effecting this separation is to relate the probability of movement from position  $E_j$  to position  $E_k$  to the number of positions at  $E_k$  currently available in the society.

The original entries, in percentage form, are the probability that a son will be in class  $E_k$  given that the father is in class  $E_j$ , written

$$P(E_k|E_j) = \frac{\text{sons, with fathers in } E_j, \text{ who enter } E_k}{\text{all sons with fathers in } E_j}.$$

This expression is then divided by the proportion,  $P(E_k)$ , of all sons who have entered occupation  $E_k$ , which gives

$$M_{jk} = \frac{P(E_k|E_j)}{P(E_k)}$$

where  $M_{jk}$ , the mobility ratio, is the transformed entry that expresses the amount of movement (or in the diagonal cells, inheritance) from  $E_j$  to  $E_k$  relative to the number of "openings" at  $E_k$  in the society (see Tables 1 and 2).



**Table 1 — Sons' occupation by fathers' occupation, Indianapolis, 1940,  
percentage distribution and mobility ratios**

SONS' OCCUPATION	FATHERS' OCCUPATION										Total	
	Professional	Semiprofessional	Proprietors, managers, etc.*	Clerical and sales	Skilled	Semiskilled	Unskilled	Protective service	Personal service	Farming	%	N
Professional	28.3% (5.1)	15.8% (2.8)	7.7% (1.4)	7.7% (1.4)	3.3% (0.6)	2.5% (0.4)	2.4% (0.4)	2.5% (0.4)	4.9% (0.9)	3.8% (0.7)	5.5	548
Semiprofessional	6.3 (2.0)	19.3 (6.2)	3.4 (1.1)	5.2 (1.7)	2.9 (0.9)	2.1 (0.7)	1.5 (0.5)	0.8 (0.3)	4.3 (1.4)	1.7 (0.5)	3.1	307
Proprietors, managers, etc.*	7.6 (1.1)	3.5 (0.5)	17.6 (2.7)	7.6 (1.2)	4.3 (0.6)	4.1 (0.6)	2.8 (0.4)	6.6 (1.0)	5.5 (0.8)	5.9 (0.9)	6.6	656
Clerical and sales	27.9 (1.3)	17.5 (0.8)	30.6 (1.4)	42.2 (1.9)	19.1 (0.9)	17.3 (0.8)	13.1 (0.6)	22.8 (1.0)	17.1 (0.8)	15.2 (0.7)	22.1	2,188
Skilled	15.4 (0.7)	23.7 (1.1)	14.3 (0.6)	15.1 (0.7)	32.3 (1.5)	18.4 (0.8)	15.4 (0.7)	17.0 (0.8)	22.6 (1.0)	23.1 (1.1)	21.9	2,163
Semiskilled	9.5 (0.4)	12.3 (0.4)	19.8 (0.7)	16.4 (0.6)	26.9 (1.0)	43.2 (1.6)	30.0 (1.1)	31.5 (1.2)	29.9 (1.1)	28.8 (1.1)	27.1	2,678
Unskilled	2.5 (0.4)	2.6 (0.4)	2.5 (0.4)	2.4 (0.3)	5.6 (0.8)	5.3 (0.8)	28.6 (4.1)	8.7 (1.3)	3.7 (0.5)	8.9 (1.3)	6.9	684
Protective service	0.8 (0.4)	1.8 (0.8)	1.6 (0.7)	1.3 (0.6)	2.1 (0.9)	2.2 (1.0)	2.4 (1.0)	8.3 (3.6)	1.8 (0.8)	3.6 (1.5)	2.3	229
Personal service	1.5 (0.4)	3.5 (1.0)	2.1 (0.6)	1.9 (0.6)	3.0 (0.9)	4.3 (1.3)	3.6 (1.1)	1.2 (0.4)	10.4 (3.1)	5.1 (1.5)	3.4	334
Farming	0.2 (0.2)	0.0 (0.0)	0.5 (0.5)	0.2 (0.2)	0.6 (0.6)	0.6 (0.6)	0.3 (0.3)	0.4 (0.4)	0.0 (0.0)	4.2 (3.9)	1.1	105
Total %	100.0	100.0	100.1	100.0	100.1	100.0	100.1	99.8	100.2	100.3	100.0	
N	474	114	1,203	1,092	2,729	1,520	720	241	164	1,635		9,892

\* Includes officials

Source: Adapted from Rogoff 1953, tables 4 and 53, pp. 48, 118.

**Table 2 — Sons' occupation by fathers' occupation, U.S. national sample, 1957,  
percentage distribution and mobility ratios**

SONS' OCCUPATION	FATHERS' OCCUPATION							Total	
	Professional	Business	White collar	Skilled manual	Semiskilled	Unskilled	Farmer	%	N
Professional	40.4% (4.8)	18.3% (2.2)	20.3% (2.4)	8.5% (1.0)	2.3% (0.3)	1.5% (0.2)	2.5% (0.3)	8.4	86
Business	19.1 (1.4)	25.8 (2.0)	17.4 (1.3)	13.6 (1.0)	6.3 (0.5)	6.1 (0.5)	11.2 (0.8)	13.2	135
White collar	12.8 (0.9)	22.5 (1.6)	24.6 (1.8)	15.6 (1.1)	17.2 (1.2)	10.6 (0.8)	8.4 (0.6)	14.0	143
Skilled manual	19.1 (0.7)	15.0 (0.6)	20.3 (0.8)	42.2 (1.6)	28.9 (1.1)	36.4 (1.4)	21.6 (0.8)	26.5	271
Semiskilled	2.1 (0.1)	12.5 (0.7)	10.1 (0.6)	14.6 (0.8)	32.8 (1.9)	27.3 (1.6)	16.5 (1.0)	17.3	177
Unskilled	4.3 (0.5)	1.7 (0.2)	5.8 (0.6)	4.5 (0.5)	10.2 (1.1)	15.2 (1.7)	13.5 (1.5)	9.1	93
Farmer	2.1 (0.2)	4.2 (0.4)	1.4 (0.1)	1.0 (0.1)	2.3 (0.2)	3.0 (0.3)	26.4 (2.3)	11.5	118
Total %	99.9	100.0	99.9	100.0	100.0	100.1	100.1	100.0	
N	47	120	69	199	128	66	394		1,023

Source: Adapted from Jackson & Crockett 1964, p. 7.

Adjustment for occupational birth rate. The chances of movement into a particular occupation are also affected by differences in the birth rate of the various occupational classes. If, for instance, doctors had very few children, then, assuming a constant size of the medical profession, the medical replacements of the next generation will tend to come more largely from the children of other occupational groups.

The effects on mobility of the occupational distribution and of differential occupational birth rates are sometimes termed *structural components* of intergenerational mobility. These are to be contrasted with the effect of parental status (and of factors correlated with it).

Interpreting mobility ratios. The calculation of  $M_{jk}$  provides a standard in terms of which mobility may be viewed as high or low. If sons were distributed in occupations on a purely random (chance) basis, then the sons of any given parental class would enter the various occupations simply in proportion to the size of that occupation in the society. In this case  $M_{jk}$  has the value of 1.0. A value of  $M_{jk}$  greater (or less) than 1.0 signifies that sons from a particular class of fathers are entering an occupation more (or less) frequently than would be expected on a purely chance basis. Thus, deviations of  $M_{jk}$  from 1.0 indicate the operation of factors associated with "father's status."

### Findings of intergenerational studies

Tables 1-4 provide illustrative intergenerational mobility findings. In Tables 1 and 2 the upper entries in each cell are percentages, and the lower, parenthetical entries are the cell values for  $M_{jk}$  (see above). These are summary tables and do not permit analysis in terms of particular age groups or

other demographic subdivisions. These and numerous other tables to be found in the literature suggest the following statements:

(1) Mobility tables uniformly show deviation from random distribution, that is, they show that filial status is statistically (and positively) dependent on parental status in varying degrees (see Tables 3 and 4). Impressive as this relationship may appear to casual inspection, it is equally apparent that the sons of most classes of fathers are distributed in substantial numbers throughout most of the status classes. Evidently, then, the status of the father permits considerable variation in the status of the son. A more precise summary statement of the over-all relationship suggested by the available studies is that probably not more than one-quarter of the variance in filial status is accounted for by parental status; and this includes the effect of some factors correlated with, but not properly included in, parental status (e.g., race).

(2) The sons are most heavily overrepresented (as compared with random expectation) in the diagonal cells, that is, in those cells representing inheritance or a continuity by the son of the parental status. This necessarily implies underrepresentation in some other cells. This underrepresentation is generally spread over a larger number of cells and is, therefore, less striking, except at times when the parental and filial statuses are in very marked contrast. It follows, then, that for those sons who enter an occupational class different from that of their fathers, *which* particular other occupational class will be entered is, generally, less dependent upon the paternal status.

(3) The above findings lend themselves to two rather different emphases. On the one hand, it is probably correct to say that only one-quarter or

Table 3 — Sons' occupational status by fathers' occupational status, Great Britain, 1949, percentage distribution

SONS' PRESENT STATUS CATEGORY	FATHERS' STATUS CATEGORY							Total	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	%	N
(1) Professional; high administrative	38.8%	10.7%	3.5%	2.1%	0.9%	0.0%	0.0%	2.9	103
(2) Managerial; executive	14.6	26.7	10.1	3.9	2.4	1.3	0.8	4.6	159
(3) Inspectional; supervisory; other nonmanual (higher grade)	20.2	22.7	18.8	11.2	7.5	4.1	3.6	9.4	330
(4) Inspectional; supervisory; other nonmanual (lower grade)	6.2	12.0	19.1	21.2	12.3	8.8	8.3	13.1	459
(5) Skilled manual; routine grades nonmanual	14.0	20.6	35.7	43.0	47.3	39.1	36.4	40.9	1,429
(6) Semiskilled manual	4.7	5.3	6.7	12.4	17.1	31.2	23.5	17.0	593
(7) Unskilled manual	1.5	2.0	6.1	6.2	12.5	15.5	27.4	12.1	424
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	
% N	129	150	345	518	1,510	458	387		3,497

Source: Adapted from Glass 1954, p. 183.



less of the over-all variance in filial status is accounted for by parental status and that consequently other factors, taken collectively, play a more important role in determining the status of the son than does parental status. At the same time, it is possible to select particular cells and quite correctly emphasize the large deviations from random expectation in these cases: for example, the considerable excess representation in professional occupations of sons of professional fathers; or the considerable deficiency among professional workers of sons of unskilled workers.

(4) The probability that the sons of a particular class of fathers,  $E_f$ , will achieve a given status level,  $E_k$ , is inversely proportional to the status "distance" between social positions  $E_f$  and  $E_k$ . Since the status "distance" is itself a function of several variables, such as education and income, this summary formulation cloaks a number of more specific relations of interest (see below).

(5) Much discussion has been devoted in recent years to two questions: (a) whether the rate of mobility has changed during the last generation or two; and (b) whether European societies show a lower rate of mobility (a greater continuity of family status from generation to generation) than the United States. Despite the number of mobility studies now available, numerous difficulties with respect to their design and comparability preclude confident answers to the foregoing questions. The following statements are therefore *tentative*.

Studies of the United States and most Western industrial societies with increasing urban sectors and considerable provision for education show in the last two generations no substantial over-all change in the tendency of sons to inherit the father's occupational class, at least when these are rather broadly defined. The available data do not suffice unequivocally to detect smaller changes that may in fact have occurred. It is evident that the decline in agricultural employment and some

skilled crafts, and the emergence of new occupations, operate to reduce occupational inheritance. But such changes have not been confined to any one generation and therefore do not necessarily produce drastic alterations in the mobility pattern. This is particularly true when social position is defined in terms of occupational prestige. The new occupations may leave unaltered the relative frequency of different status levels in the society and the distribution of the labor force among them. The increased access to educational facilities and the growth of large-scale enterprises also suggest that the data should show a decreasing dependence of filial status on parental status; but it may well be that more limited access to education in earlier years was offset by the correspondingly weaker emphasis on formal educational requirements and a greater reliance on apprenticeship and learning "on the job."

The United States has often been viewed as a society in which individual effort and merit are rewarded more substantially than in European societies and where family background counts for less than elsewhere in the distribution of status positions. Available studies, on the other hand, show no striking differences between the amount of mobility in the United States and western European societies (compare Tables 1 and 2 with Tables 3 and 4). However, most international comparisons have, in the search for comparability, reduced the data of individual studies to a least common denominator which denudes them of much of their value, and it is dubious whether, even so, comparability has been achieved. Nonetheless, there is one characteristic of U.S. society, absent in any similar degree from European societies, that may account for the difference between common assumptions and research findings; namely, a large Negro population subject to severe handicaps in the selection process. The position of the Negro in U.S. society adds to the correlation between pa-

Table 4 — Sons' social status at age 30 by fathers' social status at age 30, Denmark, 1954

SONS' SOCIAL STATUS	FATHERS' SOCIAL STATUS					Total	
	(1) (Highest)	(2)	(3)	(4)	(5) (Lowest)	%	N
(1) (Highest)	41.9%	5.1%	2.1%	.9%	.6%	2.3	45
(2)	25.8	30.8	12.8	3.9	2.2	9.4	185
(3)	16.1	30.3	36.0	20.7	13.4	24.1	474
(4)	12.9	23.9	33.3	45.2	36.9	36.9	726
(5) (Lowest)	3.2	9.8	15.8	29.3	46.9	27.4	540
Total %	99.9	99.9	100.0	100.0	100.0	100.1	
N	31	234	531	675	499		1,970

Source: Adapted from Svalastoga 1959, p. 324.

rental and filial status in the lower strata of the status hierarchy. When the upward mobility rate of unskilled white workers is considered separately, it is appreciably higher than the upward mobility rate for the total unskilled group. The assumption that the United States has higher mobility rates than European countries may, then, rest in part on a disregard of a sizable sector of the society.

Finally, if, as suggested above, parental status accounts at most for only one-quarter of the variance in the distribution of filial status, considerable differences could exist in the operation of different social-selection systems without necessitating correspondingly great differences in the specific effect of parental status.

### Determinants of social mobility

Although the mobility table has been the principal product of many mobility studies, it leaves unanswered many questions central to the theory of social selection and social mobility. The "variable" parental occupation or status embraces, or rather conceals, a host of more specific influences. Research on social mobility is now increasingly directed toward untangling the roles of these more specific variables and shows a corresponding decrease of interest in simply adding new mobility tables to those already available. Further, with respect to the question of how people get sorted into different occupations or status levels, the mobility table can provide at best only a very partial answer, that is, an answer in terms of the statistical dependence of filial status on parental status. But there are clearly many other factors that determine occupational and status selection.

**Father's occupation.** Parental occupation or status is related to the probability of filial entry into an occupation in two principal ways: (a) the father's occupational status may be correlated with a variety of filial attributes, such as education, intelligence, and race, that affect the son's occupational locus; (b) the parental occupational status may affect filial occupational locus more directly: the father's occupational experience may influence his son's occupational interests and may provide him with special knowledge, experience, incentives, and opportunities for access to it or other occupations.

**Educational level.** A substantial portion—substantial, relative to other variables—of the variation in status is accounted for by variations in educational level. Educational level is, of course, in considerable measure dependent upon the status level of parents. This dependence is lessened by increases in the society's investment in educational

facilities and the degree to which these make educational opportunities available without respect to social origin. To the extent that this occurs, the relation of education to status achievement is freed from an intermediate dependence on parental status. Sons of similar parental status show variations in educational levels, and the effect of these variations on status certainly cannot be ascribed to parental status. However, the extension of educational opportunities may at the same time reduce the correlation between status achievement and education. The more widespread a certain level of education becomes (for example, primary or secondary education), the less will variations in status, especially in the lower reaches of the status hierarchy, depend on variations in educational level—provided, of course, that increased access to education does not change the status significance of the occupational groups.

**Intelligence and mobility.** Part of the effect of education on status achievement is due to the correlation of education with intelligence. In the process of status selection, variations in intelligence operate to influence the achieved level of social status, both by leading to variations in educational level and (for persons of similar educational level) by facilitating the advancement of those of greater intelligence. Certain educational attainments have, however, become such decisive prerequisites for entry into many occupational positions that high intelligence without the added attainment of a corresponding education is unable to produce its full potential effect. Even were educational opportunities commensurate with intellectual capabilities, the correlation of intelligence with status achievement would be limited by the dependence of achievement on motivational and other personality characteristics. An equally important limitation is the fact that, except in a certain gross sense, intelligence is not an overriding criterion in occupational selection and advancement.

Intelligence as customarily measured depends in part, in its turn, on environmental circumstances. But the present state of investigation also leads one to conclude that, even as currently measured, intelligence has a major genetic component. Consequently, part of the dependence of filial on parental status is due to two sources of correlation between parental and filial intelligence—social and genetic.

Recent studies have been effective in demonstrating the existence of complex interactions between parental status, education, and intelligence and in establishing that each of these variables plays a significant role in determining filial status,



both independently and by mediating the influence of the others. But tested models do not now exist that enable one to state quantitatively the probable change in status ensuing from an increment in one of the variables (occurring at a specified stage of an individual's life), given the values of the other variables. The introduction of a genetic component of intelligence complicates the task of building such a model but at the same time appears indispensable if such models are to be used to derive a picture of the evolution of the mobility or social-selection system over time.

**Discrimination.** In some societies the existence of sizable racial or other groups subject to various modes of discrimination increases the dependence of filial status on parental status. The common obstacle shared by father and son tends to show up as a correlation between their status positions. Since status position as defined by mobility studies does not usually include the criterion of race, this necessarily inflates the degree of dependence of filial status on paternal status.

**Other social handicaps.** Relatively little attention has been paid in mobility studies to the role of special deficits that are not severe enough to exclude persons from the labor force and yet act as powerful handicaps to occupational achievement. High-grade mental deficiency, physical disabilities, chronic disease, mental disorders, alcoholism, etc., taken together, have a sufficiently high incidence and a sufficiently decided effect on occupational achievement to influence mobility tables. It is possible that a substantial part of the cases of extreme downward mobility can be accounted for in this way.

**A neglected source of downward mobility.** A further source of downward mobility may be viewed either as the result of a bias in the design of mobility studies or as an intrinsic feature of the parent-son status relationship. Mobility studies generally draw a sample of the gainfully occupied within the age range chosen for the investigation. The sample usually includes unmarried subjects and married subjects with and without children. The father sample, arrived at through the son sample, will, however, necessarily include only persons of the preceding generation who have at least one son and are therefore (in most cases) married. Since marriage and sometimes fertility are associated with greater occupational stability, mental and physical health, and general achievement, the fathers of the subjects will represent a special sample of their generation. If this is viewed as a bias in comparing the status of fathers and sons,

it could be overcome by choosing the subject (son) sample only from those in the labor force who are married and have at least one child. Since, however, all sons *must* come from the special (father) sample of the preceding generation, it is more useful to view the differential character of the father and son groups as one source of downward mobility. Given the biological and social significance of marriage and, perhaps, fertility, fathers are, other things being equal, superior as a group to an *unselected* sample of their sons. Consequently, one should expect that for this reason a number of the sons will arrive at status positions inferior to those of their fathers.

### How important is parental status?

The variables cited above account for a large measure of the filial status distribution. Does this mean then that the emphasis on the derivation of filial status from parental status has been misplaced? In part, the answer is certainly, Yes. But there is, nonetheless, a danger that the reduction of parental status to a series of more specific and often independently operating factors may lead to a neglect of those sources of influence in the parental occupational situation that exercise a direct influence on the son's occupational destiny by giving him special knowledge, incentives, and opportunities with respect to particular occupations. In dealing with sons who are professionals (or even more so, who are white-collar workers) it is fairly easy to account for their occupational status without making an appeal to the *specific* occupational locus of their parents. But if we are required to predict, not which sons will become professional or white-collar, but rather which sons will become doctors or cabinetmakers, then whether the parent is or is not a doctor or a cabinetmaker is still of considerable importance.

Behind the propensity of some classes of fathers to produce sons who follow in their occupational footsteps appear to lie certain relationships that, however, can be stated only very tentatively. Sons seem to be more likely to pursue their fathers' occupations under certain conditions: (1) if the fathers are self-employed; (2) if the self-employed fathers utilize a substantial capital in the pursuit of their self-employed occupations; (3) if entry into the father's occupation is regulated by licensing, examinations, union control, apprenticeship, or other obstacles that the parental status may aid the son to overcome; (4) if the parental occupation requires special training or education. Naturally, these relationships operate more effectively if the

occupation involved provides satisfactory rewards relative to alternatives open to the sons.

There are numerous other individual and family attributes that affect the probability that a child will attain a given status position. The number of children in the family and the birth order of a child may in some institutional settings be particularly important. The motivations and aspirations of young people in different sectors of society obviously play an important role in determining the manner in which various individual and family assets and handicaps exercise their influence.

### Social mobility and social structure

Finally, at least a brief word must be said about the relation of social selection and social mobility to the principal institutional structures of society. The research to which we owe the many mobility tables now available has been mostly pursued in Western industrialized societies with large urban sectors. This has made it easier to disregard the role of major institutional differences in the formation of social-selection systems. Current studies have tended to confine their attention more particularly to demographic and technological changes and to the role of educational institutions and the practices which affect access to them. Although there is a considerable literature on employment procedures, the study of social mobility has not adequately taken account of the fact that the occupational distribution process is a dual process, in which two sets of preferences and decisions, those of the employee and those of the employer or manager, confront each other, and that the distributive outcome is affected by the supply and demand of various qualities.

There are other major institutional features that, because of their relative stability, best reveal their relation to social mobility when it is studied over quite long time spans. Even the most stable political, juridical, and economic institutions of a society are of capital importance for the mobility process. Thus, the occupational selection process in Western society is decisively influenced by the nature of the labor contract, by laws relating to freedom of movement, and, ultimately, by the distribution of political power, together with the political and social sentiments associated with this distribution. As increasingly complex models of the social-selection process are developed, it will become necessary to specify more explicitly and exactly the institutional environment to which the model has application. Perhaps, too, it will become possible to relate parameter values to changes in

the institutional environment and thus to unify the interests of quantitative research and comparative historical inquiry.

HERBERT GOLDHAMER

[See also STRATIFICATION, SOCIAL. Other relevant material may be found in CASTE; MARKOV CHAINS; OCCUPATIONS AND CAREERS.]

### BIBLIOGRAPHY

*Bibliographical resources for the study of social mobility are excellent. For the earlier literature Sorokin 1927-1941 is still valuable. Lipset & Bendix 1959 provides an extensive listing of relevant literature in the course of a broad survey of occupational mobility, as well as a presentation of the authors' own research. Consult also Mack et al. 1957 and Miller 1960. These sources can be brought up to date by consulting Sociological Abstracts.*

*For general surveys of social mobility, see Sorokin 1927-1941; Lipset & Bendix 1959; Barber 1957; Bendix & Lipset 1953. Studies of social mobility with national, regional, and special occupational samples are too numerous to list in detail. Representative studies for the United States are Taussig & Joslyn 1932; Warner & Abegglen 1955; Rogoff 1953; Jackson & Crockett 1964; Jaffe & Carleton 1954. For Great Britain, Ginsberg 1932; Glass 1954. For France, Bresard 1950; Desabie 1956. For the Scandinavian countries, Geiger 1951; Carlsson 1958; Svalastoga 1959. For Germany, Bavaria, Statistisches Landesamt 1930; Janowitz 1958; Bolte 1959. For Japan, Nishira 1957. For Italy, Chessa 1912; Livi 1950. For the USSR, Inkeles 1950. For other countries, and for additional material on the foregoing countries, consult World Congress of Sociology. On the relation of education and intelligence to social selection and social mobility, see, in addition to the major sources already cited, Anderson et al. 1952; Conway 1958; Halsey et al. 1961; Duncan & Hodge 1963.*

*On the technical and methodological problems of studying social mobility, see especially Carlsson 1958; Svalastoga 1959; Duncan & Hodge 1963. A developing literature on mathematical models of the mobility process includes Prajs 1955; Blumen et al. 1955; Beshers & Reiter 1963; White 1963. A number of topics in social mobility (career mobility, three-generational mobility, social mobility of women, etc.) receive scant or no attention in this article. However, the reader will have no difficulty in finding material on these topics if he consults the more general works in the literature cited here.*

ANDERSON, C. ARNOLD; BROWN, J. C.; and BOWMAN, M. J. 1952 Intelligence and Occupational Mobility. *Journal of Political Economy* 60:218-239.

BARBER, BERNARD 1957 *Social Stratification: A Comparative Analysis of Structure and Process*. New York: Harcourt.

BAVARIA, STATISTISCHES LANDESAMT 1930 *Sozialer Aufstieg und Abstieg im deutschen Volk*. Beiträge zur Statistik Bayerns, Vol. 117. Munich: Lindauer.

BENDIX, REINHARD; and LIPSET, SEYMOUR M. (editors) (1953) 1966 *Class, Status, and Power: Social Stratification in Comparative Perspective*. 2d ed. New York: Free Press.

BESHERS, JAMES M.; and REITER, STANLEY 1963 Social Status and Social Change. *Behavioral Science* 8:1-13.

BLUMEN, ISADORE; KOGAN, M.; and MCCARTHY, P. J. 1955 *The Industrial Mobility of Labor as a Probability Process*. Ithaca, N.Y.: Cornell Univ. Press.

BOLTE, KARL M. 1959 *Sozialer Aufstieg und Abstieg: Eine Untersuchung über Berufsprestige und Berufsmobilität*. Stuttgart (Germany): Enke.



- BRESARD, MARCEL 1950 *Mobilité sociale et dimension de la famille*. Population 5:533-566.
- CARLSSON, GÖSTA 1958 *Social Mobility and Class Structure*. Lund (Sweden): Gleerup.
- CHESBA, FEDERICO 1912 *La trasmissione ereditaria delle professioni*. Turin (Italy): Bocca.
- CONWAY, J. 1958 The Inheritance of Intelligence and Its Social Implications. *British Journal of Statistical Psychology* 11:171-190.
- DESABIE, J. 1956 La mobilité sociale en France. *Bulletin d'information* 1:25-63.
- DUNCAN, OTIS D.; and HODGE, R. W. 1963 Education and Occupational Mobility: A Regression Analysis. *American Journal of Sociology* 68:629-644.
- GEIGER, THEODOR 1951 *Soziale Umschichtungen in einer dänischen Mittelstadt*. Copenhagen: Munksgaard.
- GINSBERG, MORRIS 1932 *Studies in Sociology*. London: Methuen. → See especially pages 160-174, "Interchange Between Social Classes."
- GLASS, DAVID V. (editor) 1954 *Social Mobility in Britain*. London: Routledge.
- HALSEY, A. H.; FLOUD, JEAN; and ANDERSON, C. ARNOLD (editors) 1961 *Education, Economy, and Society: A Reader in the Sociology of Education*. New York: Free Press.
- INKELES, ALEX 1950 Social Stratification and Mobility in the Soviet Union: 1940-1950. *American Sociological Review* 15:465-479.
- JACKSON, ELTON F.; and CROCKETT, HARRY J. JR. 1964 Occupational Mobility in the United States: A Point Estimate and Trend Comparison. *American Sociological Review* 29:5-15. → Provides a table based on a U.S. national sample, and reviews earlier national sample studies.
- JAFFE, ABRAHAM J.; and CARLETON, R. O. 1954 *Occupational Mobility in the United States: 1930-1960*. New York: King's Crown Press. → A cohort analysis of the American working force, with projections of mobility trends.
- JANOWITZ, MORRIS 1958 Social Stratification and Mobility in West Germany. *American Journal of Sociology* 64:6-24.
- LIPSET, SEYMOUR M.; and BENDIX, REINHARD 1959 *Social Mobility in Industrial Society*. Berkeley: Univ. of California Press.
- LIVI, LIVIO 1950 Sur la mesure de la mobilité sociale: Résultats d'un sondage effectué sur la population italienne. Population 5:65-76.
- MACE, RAYMOND W.; FREEMAN, L.; and YELLIN, S. 1957 *Social Mobility: Thirty Years of Research and Theory: An Annotated Bibliography*. Syracuse Univ. Press.
- MILLER, S. M. 1960 Comparative Social Mobility: A Trend Report and Bibliography. *Current Sociology* 9, no. 1:1-89.
- NISHIRA, SIGEKI 1957 Cross-national Comparative Study on Social Stratification and Social Mobility. Institute of Statistical Mathematics, Tokyo, *Annals* 8:181-191.
- PRAIS, S. J. 1955 Measuring Social Mobility. *Journal of the Royal Statistical Society Series A* 118:56-66.
- ROGOFF, NATALIE 1953 *Recent Trends in Occupational Mobility*. Glencoe, Ill.: Free Press.
- Sociological Abstracts*. → Published since 1952. Consult recent entries under "Social Stratification"; "Sociology of Occupations and Professions"; and "Sociology of Education."
- SOROKIN, PITIRIM A. (1927-1941) 1959 *Social and Cultural Mobility*. Glencoe, Ill.: Free Press. → Reprints *Social Mobility* and Chapter 5 from Volume 4 of *Social and Cultural Dynamics*.
- SYVALASTOGA, KAARE 1959 *Prestige, Class, and Mobility*. Copenhagen: Gyldendal.
- TAUSSIG, FRANK W.; and JOSLYN, CARL S. 1932 *American Business Leaders: A Study in Social Origins and Social Stratification*. New York: Macmillan.
- WARNER, W. LLOYD; and ABEGLLEN, JAMES C. 1955 *Occupational Mobility in American Business and Industry, 1929-1952*. Minneapolis: Univ. of Minnesota Press.
- WHITE, HARRISON C. 1963 Cause and Effect in Social Mobility Tables. *Behavioral Science* 8:14-27.
- WILENSKY, HAROLD L. 1966 Measures and Effects of Social Mobility. Pages 98-140 in Neil J. Smelser and Seymour M. Lipset (editors), *Social Structure and Mobility in Economic Development*. Chicago: Aldine.
- WORLD CONGRESS OF SOCIOLOGY Transactions. → Published since 1954. Transactions of the first congress, held in 1950, were not published.

## SOCIAL MOVEMENTS

The articles under this heading survey and analyze the general features of social movements and describe various approaches to their study. Specific movements and types of movements are discussed in ANARCHISM; CIVIL DISOBEDIENCE; COMMUNISM, article on THE INTERNATIONAL MOVEMENT; CONSERVATION; FALANGISM; FASCISM; IMPERIALISM; LATIN AMERICAN POLITICAL MOVEMENTS; MILITARISM; MILLENARISM; NATIONAL SOCIALISM; NATIONALISM; NATIVISM AND REVIVALISM; PACIFISM; PAN MOVEMENTS; SOCIALISM; SYNDICALISM; ZIONISM. Related modes of organization are discussed in INTEREST GROUPS; PARTIES, POLITICAL; SECTS AND CULTS; VOLUNTARY ASSOCIATIONS; and in many of the articles listed in the guide under RELIGION. For the relationship of social movements to ideology, see IDEOLOGY; INTERNATIONAL RELATIONS, article on IDEOLOGICAL ASPECTS; RADICALISM; REVOLUTION; UTOPIANISM. The impact of social movements on the social structure is reviewed in COLLECTIVE BEHAVIOR; MASS PHENOMENA; MASS SOCIETY; ORGANIZATIONS, article on ORGANIZATIONAL GOALS; POLITICAL SOCIOLOGY.

- I. TYPES AND FUNCTIONS OF SOCIAL MOVEMENTS
- II. THE STUDY OF SOCIAL MOVEMENTS

Rudolf Heberle

Joseph R. Gusfield

### I

#### TYPES AND FUNCTIONS OF SOCIAL MOVEMENTS

The term "social movement" or its equivalent in other Western languages is being used to denote a wide variety of collective attempts to bring about a change in certain social institutions or to create

an entirely new order. Sometimes the term is used in distinction from *religious* or *political* movements and from movements among particular groups, for example, the women's movement or the youth movement. As all of these movements occur in society and tend to affect, directly or indirectly, the social order, it would be permissible to apply the term social movement to all of them. However, when the term first came into use, early in the nineteenth century, it had a more specific meaning: the social movement meant the movement of the new industrial working class, with its socialistic, communistic, and anarchistic tendencies. A German scholar, Lorenz von Stein, was one of the first to recognize that the real political significance of socialism and communism lay not in their value as forms of social thought but in the fact that they gave expression and direction to the strivings of the industrial proletariat toward a new social order which would abolish economic exploitation and give the workers a chance to achieve full personality development.

On the continent of Europe, the identification of the social movement with the labor movement lasted until the second decade of this century. Today this narrow definition of the concept is no longer possible in view of peasants' and farmers' movements, of Fascism and National Socialism, and of the independence movements in colonial countries—to name only the most important instances.

It is advisable, however, to distinguish between movements which, because of their limited goals, never attract more than small groups of people and those which, aiming at comprehensive and fundamental changes in the social order, become true mass movements of historical significance. The latter will be referred to as social movements in the strict or "classical" sense. The former are sometimes referred to as "protest movements," but this term may also be used to designate mass movements which seek to redress grievances of certain groups (for example, the Negro movement) as well as movements opposing social (or political) change (for example, the Ku Klux Klan or the White Citizens' Councils in the southern United States).

Although it is sometimes convenient to distinguish between social and political movements, it should be noted that *all* movements have political implications even if their members do not strive at political power.

Social movements are a specific kind of concerted-action groups; they last longer and are more integrated than mobs, masses, and crowds and

yet are not organized like political clubs and other associations. A social movement may, however, be comprised of organized groups without having one over-all formal organization (for example, the labor movement, which comprises trade unions, political parties, consumer cooperatives, and many other organizations). Group consciousness, that is, a sense of belonging and of solidarity among the members of a group, is essential for a social movement, although empirically it occurs in various degrees. This consciousness is generated through active participation and may assume various sociopsychological characteristics. By this criterion social movements are distinguished from "social trends," which are often referred to as movements and are the result of similar but uncoordinated actions of many individuals (for example, the suburban movement, fads, and fashions).

The classical concept of social movement implies the creation of an entirely new socioeconomic and political order, especially as concerns the institutions of property and the distribution of power. To justify these aims, all major social movements develop a more or less elaborate, more or less consistent set of ideas which its members must accept more or less uncritically, as members of a religious group would accept a creed. From these "ideologies," or "constitutive ideas," are derived "action programs" of a more changing nature. Social movements tend to spread beyond the boundaries of states or national societies and to extend over the entire area of a civilization, or even beyond, as far as the social order that is their target reaches.

"Protest movements," as the term is used in this discussion, are, as a rule, limited in spatial expansion, being mostly of local, regional, or national character, for example, many "radical" farmers' and peasants' movements. A formally organized protest movement represents one kind of "pressure group." Transformation of a protest movement into a genuine social movement is possible; the early labor movement showed many traits of a protest movement, and labor unions often act as pressure groups. Neither protest movements nor pressure groups develop, as a rule, a comprehensive political action program or an elaborate ideology. While the relations between social movements and political parties will be discussed later, it should be pointed out that political parties are not necessarily differentiated by ideologies.

In spite of their different and often antagonistic aims and ideologies, certain social movements show similarities in structure, tactics, and other



formal aspects, while other movements, although similar in their ideologies, are significantly different in other respects. While the earlier studies of social movements have concentrated on the ideas, more recent sociological research has given equal attention to the structural and psychological aspects.

### Ideas

Movements aiming consciously at a radically different social order, a "change from the roots," are possible only when the social order is seen not as a divine creation but as a work of man, subject to man's will. Movements of this kind are concomitant with the secularization of thought. This is why such movements have occurred in the West only since the eighteenth century and in the East quite recently as a consequence of cultural contact with the West. Earlier revolts and disturbances among the lower social strata typically aimed at improving their social position without attacking the social order in its foundations. Radical movements in the earlier periods tended to assume the character of millenarian religious or quasi-religious sects. Similar movements still occur in regions where the transition to modern society is lagging. These movements were, as a rule, defeated unless they adopted the organization, tactics, and ideology of modern movements.

In the attempt to justify their aims, modern social movements typically resort to abstract principles concerning the nature of man, his destination, and his natural rights in combination with a critique of the existing economic, political, and cultural institutions. The ideas of liberty and equality are common to all major social movements, sometimes in combination with the idea of national unity and independence.

The most typical forms which the philosophy of a movement assumes are either a detailed, rational plan for a new society—a utopia—or, as in the case of Marxism, a theory of history which predicts the inevitable coming of the new society without revealing its form in detail. Characteristic of these thought systems is their inherent logical consistency, their reliance on monocausal explanations of major social problems, and their tendency to believe that changes in social institutions will bring final solutions to all human problems.

The defenders of the existing order will create counterarguments, which become the constitutive ideas (or ideology in the narrow sense) of a countermovement. Typically, these countermovements incorporate some of the ideas of their adversaries but otherwise use any argument that may serve

in defense of their position, regardless of resulting logical contradictions. Some have made effective use of symbols (the Third Reich) in order to achieve solidarity among adherents. Either type of belief system may be regarded as an expression of the collective will of the social groups in which they are developed and accepted.

While the methods of analyzing belief systems of social movements have been refined by the "sociology of knowledge," it is important to recognize the element of *volition* in the ideologies, since this is what makes them socially effective. The belief in a set of constitutive ideas binds the members of a movement together and gives them the *élan* needed for the persistent pursuit of the movement's aim. In some cases this belief can assume quasi-religious forms. If this happens, not only the ideas of antagonists but also any deviation from the orthodox ideology are branded as a heresy. [See IDEOLOGY.]

### Social-psychological characteristics

The motivations of individuals in joining a social movement may range from rational belief in the movement's aim (value-rational orientation) to pure opportunism. Quite often the decision to join is more emotional than rational—for example, when a person is "converted" by some experience which arouses his sense of justice—and in certain types of movements the mass of supporters are attracted by the personal charm of a leader rather than oriented toward an elaborate belief system or a definite action program (emotional-affectual orientation). When a movement has existed for a long time, it may, in certain families or occupational groups or local communities, become traditional to belong, as in the case of traditional socialistic orientation in many European working-class families. The example set by kinsmen, neighbors, or friends may give the incentive to join. [See PERSONALITY, POLITICAL.]

One can distinguish the following sociopsychological types of movements according to the prevailing motivation of their members:

(1) The value-rational "spiritual community" or "fellowship" of believers in the truth of the constitutive ideas and in the practical aims of a movement.

(2) The emotional-affectual "following" of a charismatic leader.

(3) The purposive-rational or utilitarian association for the pursuit of individual interests.

Combinations of these types are frequent, and transitions from one type to another may occur during the life cycle of a movement. What started

as a community of believers in a "cause" may develop into a "band wagon," attracting socially or politically ambitious opportunists. In some revolutionary movements the devotion to the cause becomes paramount to the degree that personal friendship and love among members and all personal, intimate relations with outsiders are discouraged—a paradox in movements which aim at the establishment of a more human, more fraternal order of society. Of course, reasons of security against police spies and traitors may contribute to these practices. On the other hand, there are cases where kinship and personal friendship contributed to the solidarity of members of a revolutionary movement.

The nature of a movement may also change as a new generation grows into it. By political generation we mean those persons of approximately equal chronological age who experienced the same crucial social and political events during the formative period of their lives, that is, roughly in their late teens and early twenties. The implication is that these experiences, among other factors, contributed to the formation of their social and political outlook on life (*Weltanschauung*). Although this concept presents serious difficulties to "operational definition," it is a most useful tool in understanding the origin and development and internal tensions and changes of social movements. [See GENERATIONS.]

Many attempts have been made to explain the origin and growth of social movements by relying only on psychic factors. These attempts have in many instances led to an overemphasis on non-rational motivations, even on pathological traits of founders and leaders as well as of followers, and have neglected the situational conditions out of which movements arise. While many early leaders of radical movements have been maladjusted personalities and while frustrated individuals are often attracted by such movements, mass adherence is gained by rational reaction to economic or other social conditions which are felt to be intolerable. Likewise, the popular notion that movements are generated by outside agitators has to be discarded; outsiders certainly accelerate the spread of a movement, but unless the local people are "ready," the agitator is likely to preach to deaf ears—or even be forced to leave.

### Organization and structure

Although social movements are by definition not corporate bodies, as action groups they need some kind of organization that enables certain persons to act as authorized spokesmen and representa-

tives. This can take the form of committees, clubs, labor unions, or political parties.

The relationship to political parties is of special importance. It may be of various patterns: (a) the party may serve as the political spearhead of the movement; (b) the movement may be represented in or exert influence through several parties; (c) a party may contain several movements—or parts of them; or (d) the movement rejects connection with any political party. Examples of (a) are the British Labour party in its earlier days and the National Socialist German Workers party (NSDAP) in relation to the broader Nazi movement; of (b) the labor movement in West Germany represented in the Social Democratic party (SPD) and the Christian Democratic Union (CDU); of (c) the major American parties; of (d) the Industrial Workers of the World (IWW) in the United States and anarchosyndicalism in general.

Political party by definition presupposes the existence of at least one other party within the body politic. However, the ideological parties which developed out of the communist and fascist movements (and some independence movements) aim at a monopoly of power and do not recognize other parties as legitimate adversaries or competitors. Since they are selective in membership and demand absolute obedience from all members, they are properly referred to as "political orders" rather than parties. These totalitarian parties aim at controlling all phases of social life through numerous affiliated organizations. [See PARTIES, POLITICAL.]

Affiliated organizations were, however, first devised by the social-democratic parties of continental Europe and later used by nearly all other parties. The primary purpose of creating women's, children's, and youth organizations together with educational, athletic, and a variety of other special-interest associations was not total control, as in the totalitarian orders, but recruiting new members and winning elections. Yet before 1933 a German socialist's life "from the cradle to the grave" could be spent in party-affiliated organizations. This indicates that for its adherents the party was more than a political institution; it was part of the broader movement in which the masses of industrial workers had found a "home"—a community of mind and spirit. The establishment of paramilitary forces by the fascist and National Socialist movements induced their opponents to follow their example.

The relations between the various organized groups constituting a movement are not always free of tension. For example, in some European countries the affiliation of trade unions with a po-



litical party has at times aroused opposition among those workers who did not share the political (socialistic) orientation of their union.

The growth of affiliated organizations together with universal franchise created problems of administration which could no longer be solved solely by reliance on part-time volunteer workers. Thus, large staffs of paid full-time workers developed in parties as well as in labor unions and other organizations within the major movements. While increasing the efficiency of the interior workings, the unwanted consequence of professionalization of the staff is often a decline in militancy. Persons who are responsible for a large organization on which people depend, will, in a critical situation, be reluctant to risk not only the lives of members but also the very existence of the organization and, eventually, its achievements. (This explains in part why the German socialist trade union leadership failed to resist Hitler in the spring of 1933.)

The power structure of social movements varies from diffusion of power to concentration of authority at the top level. Supreme authority may be institutional, that is, derived from the office to which a person has been promoted by legal procedure, or it may be charismatic, that is, derived from the belief in the extraordinary, quasi-superhuman powers of a particular person who, in turn, is motivated by the belief in his singular gifts, his call to leadership, and his political "mission." The founders and early leaders of social movements often come close to this type. However, genuine charisma is not to be confused with popularity of a leader.

In the fascist and Nazi movements, charismatic leadership was accepted as a principle and developed with all its characteristics and consequences, while Marxism-Leninism rejects the "cult of personality." This difference, obscured by labeling communist and fascist regimes as "dictatorships," is important for an adequate understanding of the structure of the two kinds of movements and of the resulting regimes. In particular, the problem of succession, which is critical in the case of charismatic leadership, is much less serious in the communist parties and regimes. [See CHARISMA; LEADERSHIP; TOTALITARIANISM.]

### Strategy and tactics

In politics, the distinction between strategy and tactics cannot be so sharply drawn as in the theory of war, but it is nevertheless important. In societies where freedom of opinion is granted, it is usually over tactics rather than strategy that a social movement will get into conflict with the government, especially if members of the social

movement engage in "direct action" (for example, sabotage, general strike, boycott, "sit-in," terror and violence) or in serious preparations for a *coup d'état*. Schisms in a social movement occur more frequently about tactics (for example, over the question of reform or revolution) than about strategy, although the more serious splits are usually those over long-range strategy (for example, the Stalinist-Trotskyite controversy). For the outsider, it is often difficult to decide whether a change in the policy of a movement is due to a change in final objectives or merely a tactical maneuver (this has been a major problem in dealing with communism).

Direct action, essentially undemocratic because it denies the opponent the chance to discuss the issue, is often resorted to when legitimate political action fails. In extreme situations the movement will culminate in a violent revolution.

The tactics and strategy of a social movement are interdependent with its ideology and its form of organization; for example, a movement aiming at revolution needs a more authoritarian organization than a movement believing in gradual reform. The choice of tactics as well as of forms of organization is in part dependent on the political system within which the movement operates and in part on the size of the movement and its influence within the political system. Therefore, the tactics of a social movement may change as it grows—they may become less revolutionary as the movement gains in influence, or they may become more aggressive as the chances of success increase.

Most social movements operate in public because publicity gives influence and increases adherence. However, secrecy is resorted to in situations where the rights of association, of assembly, and of freedom of opinion are denied or where members of a particular movement are subject to special prohibitive legislation and to persecution. The labor movement in its early phases was largely forced to secrecy. The consequence was—as it has been in similar cases—the splitting up of the larger movement into many more or less conspiratorial sect-like groups.

In political as in military action and in business, success comes to the innovator. The rise to power and the foreign-policy achievements of the Fascists and Nazis were, in part, due to the fact that they did not play according to the rules, while their antagonists at home and abroad expected them to. The same may be said, with modifications, of the communist movement: frequent tactical changes tend to confuse the adversaries. Mao Tse-tung succeeded because he deviated from orthodox Leninist strategy and tactics.

Radical revolutionary and counterrevolutionary movements are capable of violating the "rules of the game" because their members do not regard their opponents as part of the political community; they conceive of all politics in terms of friend-foe relations, where no holds are barred. This explains the use of terror before and after the seizure of power and the paradox that men who intend to create a better world for man are capable of sacrificing millions of human beings in the process.

### Class and ethnic movements

Two kinds of social movements have been of historical importance in recent times: class movements and movements of ethnic groups.

Examples of class movements are those of the middle class versus the nobility, peasants versus landlords workers against employers, farmers against middlemen, petty bourgeoisie against big business, and, more vaguely speaking, the poor against the rich. The supporting classes are typically those which, while gaining in economic significance and general socioeconomic achievements, feel economically exploited and politically suppressed. Some movements, particularly counter-movements and protest movements, originate among members of classes which are declining in socioeconomic significance. Thus, the European labor movement started among craftsmen who resented the loss of economic independence and among skilled industrial workers who represented the economic and intellectual elite of the proletariat.

A distinction should be made between peasant movements and farmers' movements. The former occur in societies where land is the property of ruling classes, who are not necessarily engaged in agriculture but who draw rent or other income in cash, or in kind, or in services from the peasants. The typical peasant movement aims at abolition of these obligations and the return of the land to its alleged or real original owners. Where peasants and landlords belong to different ethnic groups, for example, in some Latin American and in many colonial countries, the conflict is particularly sharp.

By way of contrast, modern farmers' movements occur typically among commercial farmers, especially in one-crop areas, where a high degree of economic insecurity exists. Unless there is a broad class of tenant farmers, the land question does not arise; the issues are mainly prices, interest rates, and taxes, and the main "targets" are the merchants, creditors, and the government.

Farmers' movements as a rule do not develop elaborate ideologies but raise concrete demands;

in this respect they come close to being mere protest movements. However, if their grievances are not alleviated, even modern farmers become susceptible to radical ideological movements (for example, German farmers during the rise of Nazism, French farmers and Poujadism). Peasant movements are likely to resort to violence. Their ideology, if one exists, may at the same time be traditionalistic and restorative. However, it is precisely in traditional areas of peasant unrest that contemporary communism has gained widespread support, especially in southern Europe and Latin America.

To say that a movement is supported by a given class does not mean that every member of the movement belongs to that particular class or that every member of the class belongs to the movement. These correlations are never perfect. Some movements are recruited mainly from uprooted or alienated members of certain classes (for example, most of the early Nazis, including Hitler, were *déclassé* lower-middle-class people). The founders, the leaders, and the framers of the belief system in class movements are often alienated members of another class.

Of particular importance in this respect is the role of the intelligentsia in furnishing leaders of revolutionary movements. Having no firm roots in their society, these men and women are susceptible to ideological beliefs which promise them a society in which their kind would find a satisfactory status.

The term "ethnic-group movement" is used to designate a variety of phenomena. The most important are (1) the movements for political independence of national minorities within the old empire states in Europe, (2) the independence movements of natives in colonial countries of Asia and Africa, (3) the movements for national unity—for example, in Germany and Italy in the nineteenth century and the Pan-Arabic movement in the twentieth century, (4) the movements of nationalities for civic and cultural equality within ethnically heterogeneous states (the Flemings in Belgium) and for superiority (the Finns in Finland).

As a rule these movements are led and supported primarily by the cultural, economic, and, in some cases, military elites who have vital interests in the attainment of the objectives.

The leaders of present-day independence movements in colonial countries are, with few exceptions, intellectuals and professional people, "marginal men" who have been subjected to Western culture and education. Their followers come by no means only or even in the first place from the lower social strata but, on the contrary and to a



large extent, from the recently developed classes of white-collar workers, civil servants, and military officers, as well as from the small and medium-sized businessmen, who find themselves handicapped by Western colonial rule and economic dominance. Additional support comes in many areas from miners, peasants, plantation laborers, and other categories of workers who have been in contact with Western systems of economy and government and who have been uprooted from the community of village or tribe. Even among primitive islanders in the south Pacific, movements have developed which are partly directed toward freedom from white dominance and are partly nourished by chiliastic expectations of enrichment through civilization, for example, the "cargo" cults [see NATIVISM AND REVIVALISM].

The Negro movement in the United States does not correspond to the typical pattern of ethnic-group movements because most Negroes do not aspire to political independence or cultural autonomy but to integration into American society and culture. Since the American Negroes are neither a national minority nor a social class, their movement does not aim at fundamental changes in the social order but rather at the realization of constitutional rights. However, achievement of the Negroes' aims would change not only local and regional customs but also certain parts of the existing legal order. The movement, therefore, is regarded by friend and foe as a "revolution." In its aims and tactics the Negro movement resembles the women's (suffragettes') movement. Like the early labor movement, it has received significant aid and leadership from outsiders.

In many instances one can observe that certain areas or regions of a country are the breeding places and strongholds of a variety of successive movements of a more or less radical character. This is true especially of rural areas where one-crop commercial farming in combination with other factors (for example, irregular rainfall) results in a high degree of economic insecurity among farmers or where a high degree of underemployment exists among agricultural workers. The depletion of forests and other secondary sources of farmers' income or their appropriation by outsiders also tends to create areas of social unrest (for example, northern Sweden and the cut-over areas in the Great Lakes states and the Deep South of the United States). In some of these areas social movements tend to find support among all social strata because all are affected; in other instances the property owners will lean toward one, the propertyless classes toward another, "extreme"

movement. Persistent centers of unrest and rebellion also occur in certain sections of metropolitan cities. The study of these phenomena constitutes the "ecology" of social movements.

### The functions of social movements

The great changes in the social order of the world which have occurred during the past two centuries are very largely the direct or indirect result of social movements. For even if a movement did not achieve all its goals, parts of its program were accepted and incorporated into the ever-changing social order. This has been the main, or "manifest," function of these movements.

As a movement grows, certain secondary, or "latent," functions may be observed: (1) the movement contributes to the formation of public opinion by providing for the discussion of social and political problems and through the eventual incorporation of some of the movement's ideas into dominant public opinion; and (2) it provides training of leaders who become part of the political elite and may eventually rise to the positions of leading statesmen. The socialist labor movements and the national independence movements have produced a large proportion of the present-day heads of government throughout the world. Labor leaders and leaders of other movements, even if they do not hold public office, also belong to the political elite in many countries, a fact which in some instances is recognized by heads of government, who may extend appropriate honors to such leaders or consult with them on political issues.

When these two functions have reached the point where the movement, after having changed or modified the social order, becomes part of it, the life cycle of the movement comes to an end—it has become an institution. This is true in the first place when a revolutionary movement is victorious, as in the Soviet Union and mainland China, where communism can no longer be called a movement but has been transformed into a regime. It is true also of the socialist labor movement and some farmers' movements in the economically advanced countries of northern and western Europe and in the United States and the British Dominions, while the countries where urgent social and economic reforms have been delayed or prevented are the present-day seedbeds of revolutionary socialistic and communistic movements.

RUDOLF HEBERLE

### BIBLIOGRAPHY

*The bibliography for this article is combined with the bibliography of the article that follows.*

## II

## THE STUDY OF SOCIAL MOVEMENTS

The analysis of social movements has constituted an amorphous and diffuse field of sociological research and theory that is sometimes conceived as part of the general field of collective behavior and sometimes as part of the study of voluntary associations. In this article, however, social movements are defined as *socially shared demands for change in some aspect of the social order*. This definition emphasizes the part played by social movements in the development of social change—an aspect that no student of them has been able to ignore.

At various times and under various conditions the legitimacy of a society's customary institutions or values may come under attack from different parts of the society. New arrangements are advocated; but the demand for change meets with resistance, and old habits and arrangements are maintained. Thus it comes about that groups face each other in some form of conflict.

A social movement, therefore, is not the unnoticed accretion of many unrecognized changes. Rather, it has the character of an explicit and conscious indictment of whole or part of the social order, together with a conscious demand for change. It also has an *ideological* component—that is, a set of ideas which specify discontents, prescribe solutions, and justify change.

## Directed and undirected movements

There is a mixture of formal association and informal, diffuse behavior encompassed in the concept of a movement. A significant distinction can be made between "directed" and "undirected" movements or segments of a movement. This distinction is similar to that made by Herbert Blumer (1946) between "general" and "specific" movements. The *directed* segment of a movement is characterized by organized and structured groups with specific programs, a formal leadership structure, definitive ideology, and stated objectives. Its followers are *members* of an organization as well as partisans to a belief. The *undirected* phase of a movement is characterized by the reshaping of perspectives, norms, and values which occurs in the interaction of persons apart from a specific associational context. The followers are *partisans* but need not be members of any association which advocates the change being studied.

For example, the directed segment of the feminist movement consisted of associations that sought to achieve goals of equal rights for women in various areas of American life; in its undirected

phase, the movement consisted of a subtle and unreported redefinition of the rights of women (in which both men and women played a part). In Ibsen's *A Doll's House*, Nora was a partisan participant in such a movement, without any affiliation to a feminist organization.

These two sides to the phenomenon of social movements often lead to diverse emphases among sociological studies. Those which focus on diffuse and undirected action are likely to emphasize those characteristics of movements which are akin to other phenomena of collective behavior and which are often such a striking feature of the early formation of movements and associations. Crowd and mob action, generalized unrest, and the structure of personal commitment loom large in many studies of the genesis of social movements (for example, see Cantril 1941). This type of study attempts to analyze the ways in which discontent is made specific in new definitions of rights and privileges, indictments of the existing order, and, finally, in programs for new institutional structures. For example, in George Rudé's (1959) analysis of public opinion among working-class Parisians on the eve of the French Revolution, the origin of their discontent is traced to anger at rising bread prices; only through a series of riots, statements, and counteraction did the protests of the workers become defined in an ideology of revolution.

Most studies of social movements have been researches into the career of an association, from stages of collective excitement to the activities of formally organized groups. In such studies the movement is identified with the goals and actions of organizations (see, for instance, Holtzman 1964; Lipset 1950; Webb & Webb 1894).

## Public policy and private persuasion

Directed and undirected segments of social movements affect each other. As movements arise, grow, and become recognized, they tend to generate public controversy. Uncommitted portions of the society may be polarized into partisan support on the one hand and resistant opposition on the other. The Negro civil rights movement in the United States during the 1960s transformed a relatively quiescent white population into resisters and supporters of civil rights. Even where organizational membership is lacking, people take positions and adjust their behavior in response to new expectations. Issues emerge where consensus had existed (Hyman & Sheatsley 1964).

Social movements may be distinguished according to whether they seek to achieve their goals



through *public policy* or *private persuasion*. A movement may work toward the goal of effective change in the rules of government or other institutions. For example, the Townsend movement sought to gain legislation affecting the welfare of the aged (Holtzman 1964), and the early Methodist sects hoped to reform the Anglican church (Niebuhr 1929). Other movements have concentrated on persuading individuals to a line of action and have ignored the use of public institutions as agents of control. Religious movements are more likely to seek converts than new legislation. Even in the American temperance movement the early stages were marked by efforts to persuade individuals to abstain (Gusfield 1963, chapter 2).

**Belief and politicization.** Although movements may, and often do, utilize both public policy and personal persuasion, it is important to recognize that they may shift their emphasis from one to the other. The ways in which movements become politicized is thus one of the important problems in the field. The religious expression of discontent may substitute for political expression (Lipset 1960, pp. 97-100). The contrary trend is also found; for instance, studies of millennial movements in Melanesia show that religious movements are capable of turning into nationalistic political rebellions (Worsley 1957).

The concept of a "social movement" is thus suggestive of people who, on the one hand, are in process of rejecting existing social values and arrangements, while, on the other, they are both striving to make converts to their way of seeing things and dealing with the resistance that their activities inevitably call forth. But while the movement is often carried by associations, it is not wholly an associational phenomenon. It is in the *system of generalized beliefs*, and in partisan commitment to such beliefs, that we find the characterizing features of a social movement. The unity and coherence of a movement, in its various stages and forms, depend on the similarity of the adherents' beliefs about the legitimacy of a new way of behaving, of their rejection of what has been, and of their demand for adoption of the new. This is what distinguishes a social movement from other types of special interest associations and from those hostile outbreaks and expressive protests which do not develop into explicit demands for a changed society.

### Social change and the social base

A major hypothesis in the field is that social movements are the products of social change. Circumstances arise in which long established relationships are no longer appropriate; the result of

this strain between old and new is discontent. One of the sociologist's tasks in analyzing a movement is to identify the particular social changes that have generated discontent and to specify their relation to the movement. For example, a movement toward establishing the language of Norwegian rural areas as the official Norwegian language was shown to be a product of the cultural chauvinism with which peasants responded to the influx of urban institutions and personnel in the provinces (Munch 1954).

**Social bases of movements.** Just as change is seldom uniform throughout a society, so too a social movement usually appeals to some segments of a society but not to all. In other words, it has a location in the social structure. For example, the Indian independence movement had a special attraction for Indian professionals who were, as a class, blocked from full professional attainments, although they had been given a British education (Misra 1961); the Poujadist movement in France attracted small businessmen and farmers (Lipset 1960, pp. 154-165); Methodism in its early stages had a singular appeal to the British working class (Niebuhr 1929, chapter 2).

A specific movement, of course, may appeal to more than one social or cultural segment. Analyses of movements often involve consideration of the problems posed by efforts to unite diverse social groups in one association. Thus, it has been shown that the American socialist movement was constantly troubled by the diversities arising from a rural, native American radicalism and the radicalism of urban, often immigrant, workers. Native American radicalism was both Populist and anti-industrial, whereas urban, working-class radicalism has sought to extend industrialization, provided that it is regulated (Bell 1952).

**Relative deprivation and social change.** There is no simple unilinear relation between the hardships experienced by a group and the development of movements toward change. The principle of *relative deprivation*, however, goes some way toward explaining the relation between perceived deprivation (or threat of deprivation) and the expression and organization of discontent. Research has shown that the absolute situation of a group is not as instrumental in generating and focusing discontent as is the *perception* of what is just, expected, and possible. Revolutions may, and often have, occurred just after revolutionary segments of the population have improved their economic position. Because of a group's rising expectations, the new situation may be experienced as even more distressing than the previous one. In some cases, fear of the loss of new gains may instill and in-

crease unrest. Furthermore, the loss of past status may be influential in generating movements which aim at restoration. This was one factor in the development of British trade unionism; early industrialization threatened to erase the line between craft and unskilled labor and thus threatened the position of the skilled worker (Webb & Webb 1894, chapter 1).

**Structural aspects.** Potential supporters of social movements must also be assessed from the standpoint of their skills and opportunities for the development of collective action. Social change may generate movements through structural changes, such as an increase in the capacities of groups for the tasks of communication, leadership, and organization. For example, colonial education acts as a training ground as well as a seedbed of discontent for anticolonial nationalist movements (McCully 1940).

**Ideological content.** The beliefs of any social movement reflect the unique situation of the social segments that make up its base. Taken together, these beliefs amount to a *paradigm of experience* by which the ideology and program of the movement appear right, just, and proper only to a particular segment of society, because it alone has undergone the experiences which could make the ideology seem both relevant and valid. This is true even though the ideology may be set forth in quite general terms. For instance, the movement for an amendment to the United States constitution prohibiting discrimination on the basis of sex has been stated in the rhetoric of equal rights for all women. This amendment has been sponsored, however, by upper-class women who would benefit from equality with husbands in property rights and has been opposed by working-class women who achieve special protection and welfare benefits under laws which limit their hours of work (Green & Melnick 1950). In this case, the rhetoric of "equal rights" has a different meaning for working women than for upper-class women.

### Career of social movements

In analyzing the relation of social movements to social changes, the sociologist, as we have seen, characteristically attempts to identify the social segments uniquely supportive and "open" to the movement, the social changes that produce both the discontent of these segments and their means for expressing it, and the relation between the ideological content of the movement and the specific social situations of the adherents and the partisans.

**Collective behavior to collective action.** However, discontent alone is far from being a sufficient cause of either protest activity or of more explicit

demands for change. "A movement has to be constructed and has to carve out a career in what is practically always an opposed, resistant, or at least indifferent world" (Blumer 1957, p. 147). The generalized unrest must become focused on specific parts of the social order. New beliefs must be created and adherents gained. People must be activated to champion a cause through the mobilization of their energies in a concerted manner. As this occurs, the movement often takes on new characteristics.

Whether or not an initial discontent will become organized into a movement is itself problematic. Hobsbawm (1959) has studied a number of what he calls "archaic" movements, in which diffuse protest did not coalesce into organized demands for change. Thus, Sicilian social banditry of the nineteenth and twentieth centuries, both expressed and was nourished by the diffuse protest of the Sicilian peasantry against the rich and the "outsider." However, such protest never achieved an ideological content or an organization through which to focus on a program of demands, tactics, and strategy.

The appearance of agitation in the midst of unrest aids in the development of new perspectives toward the social order and gives meaning to the social changes producing discontent. The development of leadership, organized actions, and ideology mobilizes discontent into an organized movement and frequently generates new characteristics owing to an initial outburst of energy. Thus, initial outbreaks of violence may well go beyond the attack on a specific and defined part of the *status quo*. In the Hungarian revolt of 1956, the urban riots were not revolutionary in origin, but they took on a revolutionary character as the movement favoring a new regime took on a more organized and permanent structure (Gross 1958, pp. 319-321). Without the appearance of a more permanent structure, such outbreaks are likely to be isolated events, as were the Harlem riots of 1943.

**Contingencies.** Events which are unanticipated and beyond the control or influence of the movement often change the constellation of resisting and supporting forces and thus strongly affect its career. For example, the inflation of the 1920s and early 1930s contributed heavily to the development of the Nazi party by increasing the economic discontent of lower-middle-class groups whose emotional ties to nationalism had already made them likely adherents of National Socialism (Fromm 1941; Lipset 1960, pp. 138-152).

**Structural conduciveness.** The structure of society affects the origin and form of movements in a variety of ways. This phenomenon has been re-



ferred to by Smelser (1962) as "structural conduciveness." Dissent may be condoned in one society but so prohibited in another that the movement must take the form of a secret society. For instance, according to the theory of Selig Perlman (1928), the unique features of the American labor movement were a product of the openness of the American class structure. In Europe, Perlman argued, workers could not find solutions to their economic discontent by moving out of the working class; by contrast, in the United States the workers' prospects for upward mobility were fairly good. It followed from this that American workers were less class-oriented and political in their union aims than European workers and more oriented to immediate issues of wages and conditions of work.

### The study of social movements

The literature of sociological theory contains a plethora of typologies of movements. Such typologies are often developed in terms of concern for the development of a theory of movements. However, few research studies of specific social movements have been concerned with the development of a theoretical framework; most studies have been motivated by the researcher's interest in particular social issues and philosophies. The field of social movements can be profitably described in terms of the historical, political, and ideological concerns which have been the focus of research since the mid-1920s.

**The problem of commitment.** The rise of communist parties in most Western nations as well as the development of German fascism has deeply affected students of political and social movements. Even the study of medieval and other millenarian religious movements has been actuated, in recent years, by a parallel interest in the mass political movements of the twentieth century (see MILLENARISM; see also Cohn 1957).

Sociological and social psychological studies of extremist movements have attempted to identify the sources of the organizational and ideological loyalty of adherents and partisans. "Fanaticism, enthusiasm, fervent hope, hatred and intolerance" often mark the advocate of revolutionary change (Hoffer 1951, p. xi). The fierceness with which the communists, fascists, and other "extremists" of left or right adhere to their positions of power is contrasted with the conditions of democratic politics. Rudolf Heberle (1951, chapter 15) has even drawn an analogy between totalitarian parties and religious disciplines by referring to the former as "political orders."

**Interpersonal relationships.** A major hypothesis in the study of the commitment process is that

loyalty to a movement is fostered through the network of interpersonal relationships built up in the process of participation. The initiate to a movement develops personal ties which support and enforce doctrinal commitments. The socially alienated person finds in a movement a solution to his problems of "belonging." Studies of defection among British, French, and American communists reveal that their behavior was influenced by dual and conflicting interpersonal loyalties, such as that of family or work team on the one hand and the party on the other (Almond 1954; Crossman 1949). In some movements, totalistic commitment is fostered by a round of life so controlled by organizational activities as to preclude interpersonal relations outside of the movement; where this happens, defection is tantamount to a complete reorientation of one's life. In more pluralistic movements, adherence to the movement does not cut the member loose from other competing and even conflicting roles.

**Psychological studies.** The discussion of the roots of "extremist" adherence to communism and fascism has also been approached from the standpoint of psychology. The research of T. W. Adorno and his associates (1950) into "authoritarian personality" has led to a widely held view that a constellation of masochistic and sadistic personality traits predispose some persons toward antidemocratic, intolerant, and authority-centered political positions and movements. In criticizing this thesis, Edward Shils (1954) has shown that right-wing, xenophobic, and nativist organizations in the United States have demonstrated a great inability to subordinate to united leadership and that diverse personality types are needed within the same organizational structure if a movement is to be effective. Similar psychological syndromes of intolerance for ambiguity among the partisans of both the extreme right and the extreme left in England and the United States have been discussed by Milton Rokeach (1960). Toch (1965) has developed a useful framework for analyzing the psychological motivations and consequences involved in membership commitment for a wide range of diverse movements.

### Bureaucratization of social movements

The development of at least a semipermanent organizational structure is often essential to the realization of the goals of a movement. However, such organization often sets in motion influences which defeat the ideals that gave birth to it. This is the paradox: that which is a needed means to an end is often the means which frustrates attainment of the end.

Disillusionment with the possibility of achieving ideals through organized movements has provided the motivation for many studies of social movements in the twentieth century. A major source of concern for this "blurring of mission" has stemmed from religious ideology: How can the ideals enunciated in religious thought and prophetic criticism be translated into effective institutional structures? Ernst Troeltsch (1912) was the first to pose this problem with some degree of theoretical precision when he made a distinction between *sect* and *church*. He maintained that as a sect develops into a more regularized and coordinated church organization (with a definitive membership, a trained clergy, and specified dogma and ritual), the initial mission and emotional drive of sectarianism is diffused in the accommodative perspective and rational character of the church. There is some confirmation of this hypothesis in various studies of American and British religious movements; new sects develop among the "disinherited" poor who find that the established churches no longer meet their needs. Such "churches of the disinherited" in turn display tendencies to become more conservative or "churchlike" (Niebuhr 1929; Pope 1942; Harrison 1959).

**Routinization and commitment.** Contemporary sociologists have often pointed out how the idealism and missionary zeal of spontaneous emotional commitment to a cause tends to be "corrupted" by the tendency of all organizations to become "ends-in-themselves." Max Weber's doctrine (1922) of the *routinization of charisma* is the leading theoretical statement of this view.

A number of studies of political and other movements have been inspired by similar concerns. Robert Michels' study (1911) of the German Social Democratic party in the early twentieth century has been the most influential work in this vein; his doctrine of the *iron law of oligarchy* emphasized the inevitability of a conservatizing leadership in organized movements. Several recent studies have also shown the tendency of movements to remain in existence beyond the realization of their original aims (Messinger 1955; Sills 1957). Implicit in this type of study is a "natural history" theory of organizations: they simply grow old like any natural organism. But more recent research suggests that the problem of institutionalization is far more complex. There would appear to be no inherent tendency either in organizations or in the sources of organizational commitment to move toward accommodation and compromise and thus to weaken the ardor of membership and the definiteness of the program. Not all unions lose their factional characteristics (Lipset et al. 1956);

not all sects become churchlike as they become regularized and stable (Wilson 1961); not all organized movements become accommodative (Gusfield 1955). The fate of a movement is dependent upon many factors, among which are the contingencies which affect resistance, as well as those which change the character of the initial adherents.

### Mass movements and mass society

The fact that highly industrialized societies seem to be especially likely to produce mass movements of various kinds has led scholars to inquire if there is not something in the very nature of mass society that encourages this tendency. Indeed, the conception of modern society as "mass society" has been one of the leading themes in the study of contemporary social movements. Exponents of this point of view hold that in contemporary industrial society traditional groups and institutions have lost control over the loyalties and behavior of the individual. The weakening of primary group attachments and the impersonal character of large-scale organization alienate men from the sources through which a democratic political process is mediated to citizens and achieves its legitimacy. Such alienated people are easily mobilized around charismatic leaders and symbolic goals. Thus, Hannah Arendt's study (1951) of European totalitarian movements traces their development to the destruction of a class structure in which group affiliation had governed behavior and attitude.

From the standpoint of this type of theory, we would expect leaders and adherents of contemporary antidemocratic movements to be drawn from segments of the population that are not well integrated into social units. William Kornhauser (1959, chapters 9-12) has shown this to be the case for a wide range of European and American totalitarian movements: not only are marginal members of society less "open" to control by elites, organizations, and primary groups, but they are also most apt to be attracted by the camaraderie of association and by the antiestablishment ideology so often found in radical movements of the right or left.

Other writers have introduced views and findings critical of mass-society theory. They maintain that organized and integrated groups are essential to effective movements and that new perspectives toward the social order are developed in the supportive context of primary and secondary group organization. According to this view, contemporary society has not shattered group structure, nor are the conditions of modern industrial society as "alienative" as the theorists of mass society have maintained. Socially alienated people, it is argued,



lack the essential social organization within which shared meanings could emerge in the first place; contemporary social movements can be expected to arise, as have other movements, among discrete social bases. For instance, S. M. Lipset has shown that the preponderance of support for German fascism, French Poujadism, and American McCarthyism has come from the lower middle classes in response to the threats of industrial development; similarly, the preponderance of communist support in Europe has come from the working classes and not from marginal, socially alienated persons (Lipset 1960, chapter 5).

### The wake of colonialism

Since the rise of various new nations in areas that were formerly European colonies, sociologists have recognized a need for the study of anticolonial nationalism and of the impact of social change on political and social movements in nonindustrialized societies. The study of nativism and revivalism under colonial domination has shown a tendency for such movements to be the precursors of movements for national independence. In general, analysis of nationalist movements has shown their relation to changes in traditional society which are caused by colonial policy (Van der Kroef 1955).

The revolt of Negroes against white domination constitutes a special chapter in the history of social movements since World War II. While racial tensions have been part of nationalist movements in many African and Asian countries, such movements have been most striking in South Africa and the United States, where their goal has been greater equality rather than national separation. However, the general movement of developing nations toward independence has had an impact even in the United States, as indicated by the appearance of a Negro nationalist movement (Essien-Udom 1962).

JOSEPH R. GUSFIELD

[Directly related are the entries COLLECTIVE BEHAVIOR; MILLENARISM; NATIVISM AND REVIVALISM; ORGANIZATIONS, article on ORGANIZATIONAL GOALS; SECTS AND CULTS. Other relevant material may be found in MASS SOCIETY; TOTALITARIANISM; VOLUNTARY ASSOCIATIONS; and in the biographies of MICHEL; PERLMAN; TROELTSCH.]

### BIBLIOGRAPHY

Systematic treatments of social movements can be found in Blumer 1957; Heberle 1951; Smelser 1962. The interested reader should also consult the bibliography to COLLECTIVE BEHAVIOR for various attempts at the comprehensive treatment of social movements. Much additional material on individual movements will be found in the

bibliographies of the articles cited in the text, especially MILLENARISM; NATIVISM AND REVIVALISM. For religious movements, see RELIGIOUS ORGANIZATION; SECTS AND CULTS.

- ADORNO, T. W. et al. 1950 *The Authoritarian Personality*. American Jewish Committee, Social Studies Series, No. 3. New York: Harper.
- ALLARDT, ERIC 1964 Social Sources of Finnish Communism: Traditional and Emerging Radicalism. *International Journal of Comparative Sociology* 5:49-72.
- ALMOND, GABRIEL A. 1954 *The Appeals of Communism*. Princeton Univ. Press.
- ARENDT, HANNAH (1951) 1958 *The Origins of Totalitarianism*. 2d enl. ed. New York: Meridian.
- BELL, DANIEL 1952 The Background and Development of Marxian Socialism in the United States. Volume I, pages 213-405 in Donald D. Egbert and Stow Persons (editors), *Socialism and American Life*. Princeton Univ. Press.
- BENDIX, REINHARD; and LIPSET, SEYMOUR M. 1957 *Political Sociology: An Essay and Bibliography*. *Current Sociology* 6:79-169.
- BLUMER, HERBERT (1946) 1951 *Collective Behavior*. Pages 167-222 in Alfred M. Lee (editor), *New Outline of the Principles of Sociology*. 2d ed., rev. New York: Barnes & Noble.
- BLUMER, HERBERT 1957 *Collective Behavior*. Pages 127-158 in Joseph B. Gittler (editor), *Review of Sociology: Analysis of a Decade*. New York: Wiley.
- BORKENAU, FRANZ (1938) 1962 *World Communism: A History of the Communist International*. Ann Arbor: Univ. of Michigan Press. → First published as *The Communist International*.
- BÖTCHER, HANS 1922 *Zur revolutionären Gewerkschaftsbewegung in Amerika, Deutschland und England*. Jena (Germany): Fischer.
- BRINTON, CRANE (1938) 1952 *The Anatomy of Revolution*. Rev. ed. Englewood Cliffs, N.J.: Prentice-Hall. → A paperback edition was published in 1957 by Vintage.
- CANTRIL, HADLEY 1941 *The Psychology of Social Movements*. New York: Wiley.
- CANTRIL, HADLEY 1958 *The Politics of Despair*. New York: Basic Books.
- COHN, NORMAN (1957) 1961 *The Pursuit of the Millennium: Revolutionary Messianism in Medieval and Reformation Europe and Its Bearing on Modern Totalitarian Movements*. 2d ed. New York: Harper.
- CROSSMAN, RICHARD (editor) (1949) 1959 *The God That Failed*. New York: Harper. → A paperback edition was published in 1963.
- ECBERT, DONALD D.; and PERSONS, STOW (editors) 1952 *Socialism and American Life*. 2 vols. Princeton Studies in American Civilization, No. 4. Princeton Univ. Press.
- EMERSON, RUPERT 1960 *From Empire to Nation: The Rise to Self-assertion of Asian and African Peoples*. Cambridge, Mass.: Harvard Univ. Press. → A paperback edition was published in 1962 by Beacon.
- ESSIEN-UDOM, ESSIEN UDOSIN 1962 *Black Nationalism: A Search for an Identity in America*. Univ. of Chicago Press.
- FROMM, ERICH (1941) 1960 *Escape From Freedom*. New York: Holt.
- GALENSON, WALTER (editor) 1952 *Comparative Labor Movements*. Englewood Cliffs, N.J.: Prentice-Hall.
- GREEN, ARNOLD; and MELNICK, ELEANOR 1950 *What Has Happened to the Feminist Movement?* Pages 277-

- 302 in Alvin W. Gouldner (editor), *Studies in Leadership*. New York: Harper.
- GRIEWANK, KARL 1955 *Der neuzeitliche Revolutionsbegriff: Entstehung und Entwicklung*. Weimar (Germany): Böhlau Nachfolger.
- GROSS, FELIKS (editor) 1948 *European Ideologies: A Survey of 20th Century Political Ideas*. New York: Philosophical Library.
- GROSS, FELIKS 1958 *The Seizure of Political Power in a Century of Revolutions*. New York: Philosophical Library.
- GULICK, CHARLES A.; OCKERT, ROY A.; and WALLACE, R. J. 1955 *History and Theories of Working-class Movements: A Select Bibliography*. Berkeley: Univ. of California, Bureau of Business and Economic Research and Institute of Industrial Relations.
- GUSFIELD, JOSEPH R. 1955 *Social Structure and Moral Reform: A Study of the Woman's Christian Temperance Union*. *American Journal of Sociology* 61:221-232.
- GUSFIELD, JOSEPH R. 1962 *Mass Society and Extremist Politics*. *American Sociological Review* 27:19-30.
- GUSFIELD, JOSEPH R. 1963 *Symbolic Crusade: Status Politics and the American Temperance Movement*. Urbana: Univ. of Illinois Press.
- HARRISON, PAUL M. 1959 *Authority and Power in the Free Church Tradition: A Social Case Study of the American Baptist Convention*. Princeton Univ. Press.
- HEBERLE, RUDOLF 1951 *Social Movements: An Introduction to Political Sociology*. New York: Appleton. → Includes a comprehensive bibliography of earlier works.
- HEBERLE, RUDOLF 1963 *Landbevölkerung und Nationalsozialismus: Eine soziologische Untersuchung der politischen Willensbildung in Schleswig-Holstein 1918 bis 1932*. Schriftenreihe der Vierteljahrshefte für Zeitgeschichte, No. 6. Stuttgart (Germany): Deutsche Verlags-Anstalt.
- HOBSEAWM, ERIC (1959) 1963 *Primitive Rebels: Studies in Archaic Forms of Social Movement in the 19th and 20th Centuries*. 2d ed. New York: Praeger.
- HOFFER, ERIC 1951 *The True Believer: Thoughts on the Nature of Mass Movements*. New York: Harper. → A paperback edition was published in 1958 by New American Library.
- HOFFMANN, STANLEY 1956 *Le mouvement ouvrier*. Paris: Colin.
- HOLTZMAN, ABRAHAM 1964 *The Townsend Movement: A Political Study*. New York: Bookman Associates.
- HYMAN, HERBERT H.; and SHEATSLEY, PAUL B. 1964 *Attitudes Toward Desegregation*. *Scientific American* 211, July:16-23.
- KORNHAUSER, WILLIAM 1959 *The Politics of Mass Society*. Glencoe, Ill.: Free Press.
- LANDAUER, CARL 1959 *European Socialism: A History of Ideas and Movements From the Industrial Revolution to Hitler's Seizure of Power*. 2 vols. Berkeley: Univ. of California Press. → Volume 1: *From the Industrial Revolution to the First World War and Its Aftermath*. Volume 2: *The Socialist Struggle Against Capitalism and Totalitarianism*.
- LIPSET, SEYMOUR M. 1950 *Agrarian Socialism; the Cooperative Commonwealth Federation in Saskatchewan: A Study in Political Sociology*. Berkeley and Los Angeles: Univ. of California Press.
- LIPSET, SEYMOUR M. 1960 *Political Man: The Social Bases of Politics*. Garden City, N.Y.: Doubleday.
- LIPSET, SEYMOUR M.; TROW, MARTIN A.; and COLEMAN, JAMES S. 1956 *Union Democracy: The Internal Politics of the International Typographical Union*. Glencoe, Ill.: Free Press. → A paperback edition was published in 1962 by Doubleday.
- MCCULLY, BRUCE T. 1940 *English Education and the Origins of Indian Nationalism*. New York: Columbia Univ. Press.
- MANNHEIM, KARL (1929-1931) 1954 *Ideology and Utopia: An Introduction to the Sociology of Knowledge*. New York: Harcourt; London: Routledge. → A paperback edition was published in 1955 by Harcourt. Part 1 was written for this volume. Parts 2-4 are a translation of *Ideologie und Utopie* (1929); Part 5 is a translation of the article "Wissenssoziologie" (1931).
- MESSINGER, SHELDON L. 1955 *Organizational Transformation: A Case Study of a Declining Social Movement*. *American Sociological Review* 20:3-10.
- MICHELIS, ROBERT (1911) 1959 *Political Parties: A Sociological Study of the Oligarchical Tendencies of Modern Democracy*. New York: Dover. → First published as *Zur Soziologie des Parteiwesens in der modernen Demokratie*. A paperback edition was published in 1962 by Collier.
- MISRA, BANKEY B. 1961 *The Indian Middle Classes: Their Growth in Modern Times*. London and New York: Oxford Univ. Press.
- MÜHLMANN, WILHELM E. (editor) (1961) 1964 *Chiliasmus und Nativismus: Studien zur Psychologie, Soziologie und historischen Kasuistik der Umsturzbewegungen*. 2d ed. Berlin: Reimer.
- MUNCH, PETER A. 1954 *The Peasant Movement in Norway: A Study in Class and Culture*. *British Journal of Sociology* 5:63-77.
- NEUMANN, SIGMUND (1942) 1965 *Permanent Revolution: Totalitarianism in the Age of International Civil War*. 2d ed. New York: Praeger. → First published as *Permanent Revolution: The Total State in a World at War*.
- NIEBUHR, H. RICHARD (1929) 1954 *The Social Sources of Denominationalism*. Hamden, Conn.: Shoe String Press. → A paperback edition was published in 1957 by World.
- PERLMAN, SELIG (1928) 1949 *A Theory of the Labor Movement*. New York: Kelley.
- POPE, LISTON 1942 *Millhands and Preachers: A Study of Gastonia*. New Haven: Yale Univ. Press; New York: Oxford Univ. Press.
- ROKEACH, MILTON 1960 *The Open and Closed Mind: Investigations Into the Nature of Belief Systems and Personality Systems*. New York: Basic Books.
- RUDÉ, GEORGE (1959) 1960 *The Crowd in the French Revolution*. Rev. ed. Oxford: Clarendon.
- SHANNON, FRED A. 1957 *American Farmers' Movements*. Princeton, N.J.: Van Nostrand.
- SHILS, EDWARD A. 1954 *Authoritarianism: "Right" and "Left."* Pages 24-49 in Richard Christie and Marie Jahoda (editors), *Studies in the Scope and Method of The Authoritarian Personality*. Glencoe, Ill.: Free Press.
- SHILS, EDWARD 1960 *The Intellectuals in the Political Development of the New States*. *World Politics* 12: 329-368.
- SILLS, DAVID L. 1957 *The Volunteers: Means and Ends in a National Organization*. Glencoe, Ill.: Free Press.
- SMELSER, NEIL J. (1962) 1963 *Theory of Collective Behavior*. New York: Free Press.



- SOMBART, WERNER (1896) 1919 *Sozialismus und soziale Bewegung*. 8th ed. Jena (Germany): Fischer. → A translation of the 6th German edition was published in 1909 as *Socialism and the Social Movement*.
- STEIN, LORENZ VON (1850) 1964 *The History of the Social Movement in France: 1789-1850*. Introduced, edited, and translated by Kaethe Mengelberg. Totowa, N.J.: Bedminster Press. → First published in German.
- TOCH, HANS 1965 *The Social Psychology of Social Movements*. Indianapolis, Ind.: Bobbs-Merrill.
- TROELTSCH, ERNST (1912) 1931 *The Social Teaching of the Christian Churches*. 2 vols. New York: Macmillan. → First published as *Die Soziallehren der Christlichen Kirchen und Gruppen*. A paperback edition was published in 1960 by Harper.
- TURNER, RALPH H.; and KILLIAN, LEWIS M. 1957 *Collective Behavior*. Englewood Cliffs, N.J.: Prentice-Hall.
- VAN DER KROEF, JUSTUS M. 1955 Social Movements and Economic Development in Indonesia. *American Journal of Economics and Sociology* 14:123-137.
- WEBB, SIDNEY; and WEBB, BEATRICE (1894) 1950 *The History of Trade Unionism*. Rev. ed. London and New York: Longmans.
- WEBER, MAX (1922) 1957 *The Theory of Social and Economic Organization*. Edited by Talcott Parsons. Glencoe, Ill.: Free Press. → First published as Part 1 of *Wirtschaft und Gesellschaft*.
- WILSON, BRYAN R. 1961 *Sects and Society: A Sociological Study of the Elim Tabernacle, Christian Science, and Christadelphians*. Berkeley: Univ. of California Press.
- WORSLEY, PETER 1957 *The Trumpet Shall Sound: A Study of "Cargo" Cults in Melanesia*. London: MacGibbon & Kee.

## SOCIAL OBSERVATION

See FIELD WORK; OBSERVATION.

## SOCIAL ORGANIZATION

See KINSHIP; SOCIAL STRUCTURE; SYSTEMS ANALYSIS.

## SOCIAL OVERHEAD CAPITAL

See CAPITAL, SOCIAL OVERHEAD.

## SOCIAL PERCEPTION

See under PERCEPTION.

## SOCIAL PHYSICS

See GEOGRAPHY, article on STATISTICAL GEOGRAPHY.

## SOCIAL PROBLEMS

Social problems are described most simply as perplexing questions about human societies proposed for solution. The distinctiveness of such questions as a separate object of sociological study rests upon their topicality, currency, and pragmatic

derivation. Social problems are part of the climate of opinion in society which centers on expressed needs for public policies and anticipated requirements for social control. Social problems study or research consists of the ordering of perspectives and social facts in relation to the ends and means of collective action.

Proceeding beyond this general statement to a more precise definition of social problems poses a complicated task of sorting out the wide diversity of views held by sociologists as to the nature of the subject matter and the perspectives from which it should be studied (Merrill 1948). These conflicting viewpoints, as well as salient misgivings shared by many as to whether social problems is a "field," or can validly be included with sociology, are in part understandable in the context of the origins and history of sociology itself.

### History of social problems approach

Concern with social problems has been singularly American or Anglo-Saxon. Antecedents can be found in the literature of socioeconomic criticism and reform which was directed at many of the consequences of commerce, industrialism, and urban growth in western Europe, particularly eighteenth-century and nineteenth-century England. The more immediate forerunners of what came to be the social problems approach emerged from writings, reports, essays, and surveys by Protestant clergymen, philanthropists, and middle-class humanitarians, in the United States as well as England, who were dedicated to a variety of social reform activities. These included prison reform, settlement work, child rescue, promotion of temperance, housing betterment, and improvement of conditions of employment of women and children; by the middle of the nineteenth century many of these had crystallized into organized actions or associations.

The roots of the intellectual orientation toward social problems as an academic subject are more precisely located in the broadly based American reform movement from which, in 1865, there issued the American Social Science Association. This represented a merger of a variety of local and regional associations, whose constituted objectives were clearly meliorative (Bernard & Bernard 1943). In large part it was responsible for the introduction of social science courses in American colleges and universities, beginning in 1865 and reaching a peak between 1885 and 1895. Many, if not most, of these courses, however titled, dealt with topics subsequently recognized as the substance of social problems courses in sociology—

with possibly somewhat greater attention paid to education and law.

The development of such courses reflected motivations of persons both within and outside the universities who were seeking to arouse and prepare students for careers of legislative reform. The courses attained quick popularity with students, many of whom were repelled by limitations of the classical or science curricula and who were fired by the social ferments of the post-Civil War period. Toward the end of the nineteenth century, when sociology began to receive formal departmental recognition in colleges and universities, many of those recruited to teach it came from backgrounds of the ministry and welfare work. The lineal ties of their versions of sociology to the older social science movement are attested by the substantial numbers of these early sociologists who were members of the National Conference of Charities and Corrections and of the American Prison Congress (Sutherland 1945).

**The scientific rationale.** Such facts strongly tempt one to the conclusion that American sociology was fathered by the study of social problems. However, this is opposed by another theme, reflecting continuity with the thought of August Comte and Herbert Spencer, holding to a scientific purpose in the study of society, which was present almost from the first in the social science movement. To the scientific emphasis in pioneer American sociology was added an antireform bias, stemming from the laissez-faire philosophy of Spencer and sounded in W. G. Sumner's strident disdain for social welfare activities. The conflict of purposes among early sociologists is epitomized by the lengths to which Sumner himself went in devising titles which would sharply distinguish his courses from the reformistic ones taught by his colleagues at Yale Divinity School.

It is generally accepted that Lester F. Ward's teleological philosophy outweighed the influences of Spencer and Sumner in the formative years of American sociology and that the notion of an "applied sociology" became or remained dominant. Yet the applied sociology of Ward (1906) was little more than an idea which took its concrete meaning from classroom teaching, student field trips to charitable institutions, and the writing of textbooks, a number of which appeared in the first decade of the twentieth century. These, as well as the ones which followed, drew heavily on factual data from a variety of sciences construed to have a bearing on the problems under discussion, roughly ordered into "causes," "effects," and "solutions." This was consistent with Spencer's concep-

tion of sociology as a synthetic, "capstone" science, but this scientific rationale probably made a virtue of necessity. The amount of "pure" sociology available for application to social problems was severely limited; early sociologists, like Yankee inventors, made do with materials at hand.

**Changing perspectives.** The passing of time saw the social problems approach in sociology lose ground as the socioeconomic characteristics of recruits to the field changed and as the need to validate sociology's status as a macrocosmic science was increasingly accepted. In the middle decades of the twentieth century sociologists turned more and more to self-conscious discussion of methodology, research design, and theory, with a growing attention to the European sociology of Max Weber and Émile Durkheim. A kind of ideological commitment to social neutrality and nonevaluative research took hold of the discipline. The gulf between social theory and the study of concrete social problems grew wide, intellectually stranding many sociologists with continuing interest in the latter (Davis 1957). Their discontent culminated in the establishment in 1952 of the Society for the Study of Social Problems, which, while affiliated with the American Sociological Association, nevertheless carefully cherishes its separate identity.

Despite the shift in perspective regarding the fundamental tasks in sociology, the flow of textbooks on social problems has continued unabated, and courses so titled continue to be taught—albeit more restively by a younger generation of sociologists trained to demand theoretical meaning in materials with which they deal. The attenuated identification of sociologists with social problems courses is perpetuated by the relatively large number of students attracted to their offerings who nevertheless do not plan to follow sociology as a course of study. Instruction in social problems courses in otherwise scientifically oriented sociology departments is given various rationales—general education, a "service function," or a means of recruitment of students to the field.

Other pressures also have made it difficult for sociologists to disengage themselves from the old ties to social problems. Government recognition of military uses of sociology during World War II, plus research support from industry, government, and private foundations after the war, drew the interest of sociologists toward applied research on problems posed by persons or agencies outside the field. The massive surge of the American Negro after 1954 toward greater equality of opportunity multiplied the contingents of sociologists at work on applied research. The looming threat of thermo-



nuclear war motivates others toward immediately useful, rather than abeyant, "scientistic" sociology. A number of highly articulate critics both within and outside sociology have inveighed against the sterility or inapplicability of much contemporary theory, directly challenging the claim that sociology can or should aspire to ethical neutrality.

### Theoretical issues

The most sweeping indictment of social problems writings appeared in an article by C. Wright Mills (1943) entitled "The Professional Ideology of Social Pathologists." With slashing phrases he more or less condemned an entire generation of social pathologists for the low conceptual level of their textbooks, the discrete and unrelated nature of their treatments of various social problems, and insertions of rural-biased value judgments in the guise of objective terminology. These observations, while undeniably cogent and pointed, were closer to stricture than to science, although they were organized around Mills's own brief for structural analysis of social problems data.

Closer study and more sanguine assessment of the field in the past make it less atheoretical than it would seem, especially if the textbooks are disregarded in favor of provocative articles by Frank (1925), Waller (1936), and Fuller (1937; 1938; compare Fuller & Myers 1941a). These writers, especially Fuller, both saw and sought to analyze social problems in a general setting of values and value conflict. Fuller's distinction between ameliorative problems and moral problems, implications of which he developed in a paper on morals and the criminal law (1942), bears insightfully on structural questions concerning the relationship of values to norms and the contingency of legal action on variations in this relationship.

Both Frank and Fuller stressed a holistic view frequently repeated by many contemporary sociologists, i.e., that situations or behavior considered to be problems, on finer analysis prove to be expressions of cherished values or institutionalized norms crucial to the operation of society. From this came Fuller's conclusion—dismal to some—that solutions to social problems may be or are impossible. This, when fitted with his later effort to demonstrate a natural history of social problems (Fuller & Myers 1941b; Lemert 1951a), closely allies his thinking with the *laissez-faire* philosophy of Sumner, but it is also akin to the political conservatism many sociologists believe to inhere in modern "system" theories of society. In sum, these critics seem to say that putting social problems into a structural context destroys both

ideological rationale and the individual motivation for reform, by showing the problems to be "necessary" consequences of a given type of value system or by making clear that values will have to be sacrificed and institutions disrupted if the problems are to be eliminated.

**Definition as an issue.** The first authors of books on social problems bothered little or not at all about definitions of social problems, uncritically drawing on fairly homogeneous convictions about the aspects of society that needed improvement or reform. Among the first attempts at definition were those of Ellwood (1915), Howerth (1913), Kelsey (1915), and Hart (1923). The prevailing definition, however, came from Case (1924), who was attracted to ideas of Thomas (1909) dealing with generic elements in the process of cultural origins. Predominant among these was an element of *attention*, defined as the subjective or reciprocal aspect of social control, which is activated by crises (Thomas [1917–1937] 1951, p. 218). These ideas led Case to propose that social problems are situations impressing a large number of competent observers as needing remedy by collective action. They became for him and many others after him sociopsychological phenomena; social problems, stated most simply, are whatever a goodly number of members of society say they are.

This definition more or less identifies sociologists with the lay populace and makes public opinion sociological opinion, with implied faith in a democratic process. Its difficulties accrue from recognition of the irrational or spurious qualities in public expressions or collective behavior, which counsels considerable discounting of public reactions or moral indignation as guides for sociological criticism of society or its institutions. Moreover, questions must be faced as to how many or what persons qualify as an acceptable panel for making judgments as to what are social problems. Many issues in modern society are articulated almost exclusively within coterie of specialists in health, medicine, welfare, correction, and education. They reflect technical interests, often couched in esoteric language, which are projected into the arena of public opinion only ephemerally or adventitiously.

Superficially, Fuller's distinction between moral and ameliorative problems seems to reconcile the older conception of social problems with the facts of technical specialization. However, the division between moral and technical problems often becomes vague or disappears, for means may become ends or ends means, depending upon the vantage point of the beholder. The older idea that social

problems could be defined by a consensus of professional and welfare experts made little headway with sociologists, largely because judgments of specialists outside their own or adjacent areas of interest can claim no greater validity than those of educated lay persons. Representative specialists often are spokesmen for organized groups, necessarily supporting vested agency values as well as conveying judgments derived from technical knowledge. Finally, it must be noted that the ordering of social problems with respect to priority or importance cannot be determined by consulting specialists who define them distributively.

### Social pathology

The subjectivism inherent in the "popular" definition of social problems runs athwart the conception of sociology as a body of knowledge which rises above common sense and is accumulated through application of special methods by observers or researchers at least relatively detached from the social facts under scrutiny. If social problems are defined in terms of such a body of knowledge, they become objective rather than subjective facts, "discoverable" from laws or generalizations about necessary conditions of social life. Despite its well-documented shortcomings, the bare idea of social pathology is more congenial to such a formulation of sociology than is that of social problems.

Social pathology was an effort to apply a biological or medical model to the analysis of problematical phenomena of society. It rests on the idea that societies or their constituent parts may develop abnormally or anomalously and that they can be described or "diagnosed" in the light of some pristine or universal criterion of normality or health. The orientation of social pathology was, however, toward man rather than society, being heavily pervaded with the nineteenth-century concern about the relevance of institutions to the perfectibility of human nature. The notion of individual adjustment figured large in discussions of social pathology, revolving about the consequences of physical illness, mental deficiency, mental disorders, alcoholism, lack of education, or incomplete socialization for the realization of life goals regarded as normal for most people. The fact that many of these conditions are indeed associated with organic pathologies or were assumed to have a hereditary foundation lent strength to the idea that social problems were external, or objective, facts. Sociological residues of this idea persist today among those who believe alcoholism and mental disorder to be diseases. In the perspective of time it may be said that so long as American society was domi-

nated by middle-class values, laissez-faire individualism, localism, and southern regionalism, the more absolutistic conception of social problems as social pathology remained tenable.

### Social problems and social disorganization

The growth of cultural relativity in sociology, infiltrating from the critical-historical themes of American anthropology, together with the general questioning of paramount American values that was generated by the great depression of the 1930s and by foreign revolutions, put an end to social pathology as a viable perspective on social problems. The need for concepts to organize thought about societies in wholesale flux and crisis was conjoined with the need to place discussion of social problems in a more comprehensive intellectual scheme that would be in keeping with the methodological aspirations of sociologists. The needs in part seemed met by restatement of conceptions of social disorganization originally set down by Thomas and Znaniecki (1918-1920) and Charles H. Cooley (1918). Many of the phenomena that had long been the subject matter of social problems or social pathology now were postulated as symptoms or products of such processes as uneven cultural development, conflict, dissensus, and dialectical change. Taken together, these processes mean social disorganization. A pivotal distinction was set up between social disorganization and personal disorganization, with the latter assumed to be functionally associated with the former. On this point most textbook writers followed the organic analogy of Cooley rather than the ideas of Thomas and Znaniecki, who saw no necessary relationship between the two.

While some sociologists have decried texts on social disorganization as being little different from those on social problems, apart from their introductory chapters, this ignores the lively development of ecological studies by sociologists at the University of Chicago and elsewhere, whose findings appeared to give statistical support for the notion that social problems are expressions of a common, underlying social process. For a while, at least, the idea of "disorganized areas" in urban communities became established sociology and a welcome new way of ordering the data in textbooks. Vice, crime, poverty, divorce, and mental disorder all became part of a zonal or area parcel explainable in terms of generic ecological growth, change, and deterioration.

Social disorganization as a subject has the look of an impressive theoretical façade which on closer analysis is disillusioning. The concept of process



on which it relies is vague at best, and the distinction between social and personal disorganization is difficult to maintain. Serious doubt has been cast on the method of ecological correlation, undermining the neat idea of disorganized areas. Careful ethnographic studies of the slum, such as W. F. Whyte's *Street Corner Society* (1943), make it undeniably clear that such areas can be or are relatively well organized. [See *DEVIANT BEHAVIOR*.]

The idea of social disorganization still has its theoretical partisans and, defined simply as human activities or failures to act which impede or block other activities, is indeed demonstrable and of heuristic value. Thus, if artillerymen fire shells that, by reason of faulty communication or function, fall on their own troops, the result is fittingly enough described as social disorganization. But if questions are asked whether the attack was part of integral tactics, or whether the associated campaign advanced over-all strategy, or whether the war should have been fought in the first place, theoretical analysis quickly moves into speculation where fact and value blur. In any but a specifiable, closed social system or subsystem with consensual goals, the formulation of social disorganization reduces to value choices of its author.

### Social problems as dysfunctions

Forms of thought from the traditions of European sociology and English anthropology offer the theoretical alternative of subsuming the data of social problems under the category of social dysfunctions. This descriptive and analytic device proceeds from assumptions that there are functional prerequisites of social life around which institutional structures operate, mutually supporting each other, meeting psychobiological needs of individuals, and contributing to an over-all integration of society. Practices or activities which run counter to functional prerequisites, which disrupt the institutional nexus, or frustrate individual needs are defined as social dysfunctions.

The difficulties of functional analysis are well known. Determining what is indispensable to maintain a specific complex of behavior and what is adventitious is made difficult by the fact that the range of cases or time series in human societies from which to generalize often is small and some kinds of events have only a few instances. Culture comes to each generation in unsorted bales or packets, and historical, comparative, or cross-cultural studies are limited in their potential for sorting out that which is functional or dysfunctional. When cultures change or undergo disruption, that which may have seemed to participants or ob-

servers to be a causal or functional association may turn out to be dispensable. The social problems of yesteryear often live only as the quaint reminiscences of today.

The persistence of ancient social problems or dysfunctional activities in such forms as crime, political corruption, gambling, or prostitution in the face of collective efforts to eradicate them is not readily explained by functional analysis. An explanation for such seeming paradoxes was proposed by Robert K. Merton (1949, chapter 1) in the form of a distinction between "manifest" and "latent" functions; R. M. Williams (1951) also dealt with this matter in his discussion of "patterned evasions of institutional norms." However, these explanations seem secondary or residual at best. They call attention to the fact that actions may have functions as well as dysfunctions. In another light, they are implicit concessions that determination of functions in a culturally diverse society depends in large degree on the particular needs, perspectives, or values adopted by the observer.

A more crucial but closely related issue in functional analysis is whether it can reveal those activities which can be established in objective terms as social problems, even though they are not necessarily subjects of popular awareness or collective action. In a general way Case (1924) heeded this question when, in defining social problems primarily as aspects of the collective mind, he nevertheless recognized that statisticians and others could describe *adverse conditions* of society; furthermore, the inclusion of a "situation" in definitions by a number of other sociologists was recognition that objective factors were a necessary part of social problems (Smith et al. 1955). Ogburn (1922), more clearly of functionalist persuasion and impressed by the discrepancies between dynamic technology and institutional adaptation, sought a totally objective definition of social problems as consequences or expressions of "cultural lag." Subsequent critics showed that this conception was not, as hoped, free from value judgments and strongly argued against its underlying dichotomy of material and nonmaterial culture.

The causal relationships between technology, culture, and moral ideas and the sequence in which they change continue to be among the great moot questions of sociology. They have grown more prominent as ruling groups in some societies seek to assist others to industrialize or raise agricultural productivity. The discovery in many instances that introduced technology does not always lead to the expected consequences has compelled students of

socioeconomic development to conclude that their judgment of what is functional and dysfunctional for others is not easily imposed for the purpose of directing change.

### Social problems as deviation

Since 1940 a sizable portion of the traditional subject matter of social problems, such as crime, delinquency, prostitution, drug addiction, and physical handicaps, has been categorized as deviance, deviation, or deviant behavior. The amoral, statistical, or descriptive implications of the terms carry a strong appeal, although they tend to acquire morally invidious connotations. Generally, deviance is defined as violations of norms, or departures from social expectancies, but beyond this minimal agreement the ideas projected for its analysis differ considerably.

One group of sociologists, following Durkheim, Parsons, and Merton, has concentrated on sources of deviation in discontinuities, anomie, or strain within the structure of a society that is assumed to be more or less an integrated system. The analysis of deviation in this theoretical context is voluntaristic, in contrast with deterministic or strictly causal versions of functionalism. Deviation originates from permutations of choice by individuals motivated by culturally given ends and confronted with means of varying accessibility. The most cogent statement or theoretical design derived from these ideas appeared in Merton's widely influential article "Social Structure and Anomie" (1938).

Critical assessment of the structural or "anomie" interpretations of deviation was slow to crystallize but was finally made by a symposium of sociologists qualified by extensive research in areas of deviation. In this volume, edited by Marshall B. Clinard (1964), they raised serious doubts as to whether Merton's effort to design an embracing theory of deviation was sufficient for the complexities of the data. The ends-means distinction is not an easy one to maintain with concrete data, and the individual motivational base of structural sociology is barren ground for the production of a theory of group-related deviation in any but reactional terms. The heavy accent on conditions of social order in works of Parsons (1951) reduces social control to a negative mechanism for repressing deviation; the recognition of deviation as a creative necessity for social change is absent from structural theories or appears only in revised afterthoughts.

Standing at considerable theoretical distance from the objectivism of current structural theories of deviation are the productions of a small nucleus

of sociologists represented by Erving Goffman (1963; 1961), Howard S. Becker (1963), and Edwin Lemert (1948; 1951b), whose perspectives are more microcosmic than macroscosmic. Characterized by the importance they assign to interactionally derived meanings in the genesis of deviation, they have dwelt upon factors present in interaction, such as labeling, stigma, self-presentation, identity conflict, and identity protection; they have also dwelt upon the more structural concepts of role, deviant career, societal reaction, and secondary deviation. These writers do not talk of social problems, but their interests in the self-definitional and constrictive impacts of social control align them with the notion that social problems are products of definition. They are little concerned with criteria for rational social control or solutions for problems of deviation, preferring to show how agencies of reform, rehabilitation, or "treatment" give form and meaning to deviation and stabilize it as secondary deviation by intervening interaction processes.

While these sociologists offer strong negative criticism of institutional controls and public policy relevant to deviance, thus far they have not produced explanations of how definitions of deviation and related policy appear as cultural phenomena, or of how and why they change. Neither have they done much theoretically to connect deviation with innovation and creativity in the processes of change. Broadly assessed, however, interactional studies of deviation show more sensitivity than systematic sociological theory to many salient features of modern society, among which must be numbered functionally derived morality, shifting ends-means relationships, pluralistic group action, segmentalized social relationships, information control, attenuated role commitment, and the operational consequences of these for norms and rules defining deviation.

### Future of social problems approach

In some respects social problems is one of the most important branches of sociology, for it provides the testing ground for predictions and the ultimate usefulness of sociology. Yet it remains a theoretically embarrassing area for many sociologists who regard its crucial questions of definition as unanswered. Discussions and research in social problems, particularly those subsumed as deviancy, now can lay claim to a closer alliance with general sociological theory than was true in the past. However, the dilemma between subjective and objective viewpoints persists, reviving with nearly every attempt to formally delimit the field.

There are growing indications that this dilemma



may be resolvable with theory and studies of values and valuation, concepts toward which sociologists and other social scientists have been pushed by the dissensus, group conflict, and resistance which seem to be the ubiquitous concomitants of swift change in modern society. Some sociologists, and more anthropologists, have sought to discover the basis of action in a general structural matrix of values delineated by age, sex, occupation, kinship, and other status attributes; others have looked for significant values in derivative, "second order" categories, such as "personality structure." However, these researches bypass the question of how values are aggregated for a pattern of action in any increasingly pluralistic society containing tangles of competing groups and associations.

One consequence of the heightened interest in values and valuation may be the evolution of a science of social action from the more amorphous area of social problems, implemented by studies of conflict resolution and short-term change and emphasizing content as well as forms of group interaction. This requires a different imagery of social problems, one less pointed to crisis and reform and more amenable to the facts of continuous change and policy revision in high-energy societies. The notion of a solution to social problems as a synoptical "best possible" choice from a number of alternatives will prove less useful than knowledge of how decisions are made and executed the "hard way." These decisions, it is to be hoped, will be seen less as "causes" of change in static situations than as strategies for intervention into ongoing processes. If value aggregation through group interaction is an important phase of such processes, it will be no less important to specify who has access to what means of social control and to locate the relevant values of power elites; for these, too, are among the "conditions" of change.

Successful linkage of "subjective" value phenomena with "objective" social structure and technology may come from theoretical extensions of human ecology, widely conceived to show how time, space, energy, and resistance to change intrude into the sociopsychological aspects of action. This necessitates acceptance of the idea that technology can change the order of choice among values, otherwise set by culture and social structure, by altering the costs of their fulfillment. This is not to say that human beings always perceive and respond economically to "costly" ways of acting. Both the definition of social problems and the organization of rational intervention are problematical phenomena to be explained. Insofar as social scientists assume responsibility for supplying

critical knowledge for rational intervention, they help to organize it. Consequently, it is doubtful that social problems can be studied from an ethically neutral position.

EDWIN M. LEMERT

[See also DEVIANT BEHAVIOR; SOCIAL CONTROL; VALUES.]

#### BIBLIOGRAPHY

- BECKER, HOWARD S. 1963 *Outsiders: Studies in the Sociology of Deviance*. New York and London: Free Press. → See especially Chapter 1, "Outsiders," and Chapter 2, "Kinds of Deviance: A Sequential Model."
- BERNARD, LUTHER L.; and BERNARD, JESSIE 1943 *Origins of American Sociology: The Social Science Movement in the United States*. New York: Crowell. → See especially pages 554 ff.
- CASE, CLARENCE M. 1924 What Is a Social Problem? *Journal of Applied Sociology* 8:268-273.
- CLINARD, MARSHALL B. (editor) 1964 *Anomie and Deviant Behavior: A Discussion and Critique*. New York: Free Press. → Includes "An Inventory of Empirical and Theoretical Studies of Anomie," by Stephen Cole and Harriet Zuckerman.
- COOLEY, CHARLES H. 1918 *Social Process*. New York: Scribner. → See especially pages 153-168, "An Organic View of Degeneration."
- DAVIS, ARTHUR K. 1957 *Social Theory and Social Problems: Fragments for a Philosophy of Social Science*. *Philosophy and Phenomenological Research* 18:190-208.
- ELLWOOD, CHARLES A. (1915) 1919 *The Social Problem: A Reconstructive Analysis*. Rev. ed. New York: Macmillan.
- FRANK, L. K. 1925 *Social Problems*. *American Journal of Sociology* 30:462-473.
- FULLER, RICHARD C. 1937 *Sociological Theory and Social Problems*. *Social Forces* 15:496-502.
- FULLER, RICHARD C. 1938 *Problem of Teaching Social Problems*. *American Journal of Sociology* 44:415-425, 433-435.
- FULLER, RICHARD C. 1942 *Morals and the Criminal Law*. *Journal of Criminal Law and Criminology* 32:624-630.
- FULLER, RICHARD C.; and MYERS, RICHARD R. 1941a *Some Aspects of a Theory of Social Problems*. *American Sociological Review* 6:24-32.
- FULLER, RICHARD C.; and MYERS, RICHARD R. 1941b *The Natural History of a Social Problem*. *American Sociological Review* 6:320-328.
- GOFFMAN, ERVING (1961) 1962 *Asylums: Essays on the Social Situation of Mental Patients and Other Inmates*. Chicago: Aldine.
- GOFFMAN, ERVING 1963 *Stigma: Notes on the Management of Spoiled Identity*. Englewood Cliffs, N.J.: Prentice-Hall.
- HART, HORNELL N. 1923 *What Is a Social Problem?* *American Journal of Sociology* 29:345-352.
- HOWERTH, IRA W. 1913 *Work and Life: A Study of the Social Problems of To-day*. New York: Sturgis & Walton. → See especially Chapter 1, "The Social Problem To-day."
- KELSEY, CARL 1915 *How Progress Causes Social Problems*. Pennsylvania, University of, *University Lectures Delivered by Members of the Faculty in the Free Public Lecture Course* [1914-1915]:9-16.

## SOCIAL PSYCHIATRY

See under PSYCHIATRY.

## SOCIAL PSYCHOLOGY

The article under this heading provides an overview of the field of social psychology. More detailed systematic and theoretical analyses may be found in FIELD THEORY; GESTALT THEORY; MARXISM; SOCIOLOGY; THINKING, article on COGNITIVE ORGANIZATION AND PROCESSES. Concepts of general and historical importance in social psychology are discussed in ATTITUDES; COMMUNICATION; COMMUNICATION, MASS; GROUPS; IDENTITY, PSYCHOSOCIAL; IMITATION; INTERACTION; LANGUAGE; MOTIVATION, article on HUMAN MOTIVATION; PERCEPTION, articles on PERSON PERCEPTION and SOCIAL PERCEPTION; PERSONALITY; PERSUASION; POLITICAL BEHAVIOR; PSYCHIATRY, article on SOCIAL PSYCHIATRY; PUBLIC OPINION; REFERENCE GROUPS; RELIGION, article on PSYCHOLOGICAL ASPECTS; ROLE; SELF CONCEPT; SOCIALIZATION; SUGGESTION; SYMPATHY AND EMPATHY. Other relevant material may be found in AGGRESSION; ALIENATION; COALITIONS; COHESION, SOCIAL; COLLECTIVE BEHAVIOR; COMPETITION; CONFORMITY; CONSENSUS; DECISION THEORY; MASS PHENOMENA; ORGANIZATIONS; PREJUDICE; STATUS, SOCIAL; STEREOTYPES. Methods and techniques used in social psychological research are treated in ACHIEVEMENT MOTIVATION; EXPERIMENTAL DESIGN, article on QUASI-EXPERIMENTAL DESIGN; INTERVIEWING, article on SOCIAL RESEARCH; OBSERVATION, article on SOCIAL OBSERVATION AND SOCIAL CASE STUDIES; PERSONALITY MEASUREMENT; PROJECTIVE METHODS; SOCIOMETRY; SURVEY ANALYSIS. For the historical background of modern social psychology, see the biographies of ALLPORT; COOLEY; DURKHEIM; HOVLAND; LE BON; LEWIN; MEAD; ROSS; SIMMEL; TARDE.

Although its intellectual origins extend back into antiquity, social psychology as a scientific discipline has only recently come of age. Our purpose in this brief survey will be to provide a rough sketch of the current status of the field by outlining the problems, approaches, and methods of contemporary research. Specific content areas are treated in more detail in other articles.

The set of problems on which social psychology has been focused, problems that have intrigued thinkers since ancient times, concerns the social nature of man and the manner in which this social nature develops. The awesomeness of whatever

- LEMERT, EDWIN M. 1948 Some Aspects of a General Theory of Sociopathic Behavior. Washington State Univ., Pullman, *Research Studies* 16:23-29.
- LEMERT, EDWIN M. 1951a Is There a Natural History of Social Problems? *American Sociological Review* 16:217-223.
- LEMERT, EDWIN M. 1951b *Social Pathology: A Systematic Approach to the Theory of Sociopathic Behavior*. New York: McGraw-Hill.
- MERRILL, FRANCIS E. 1948 The Study of Social Problems. *American Sociological Review* 13:251-259. → For divergent views see the accompanying "Discussions" by Ernest R. Mower and Stuart A. Queen on pages 260-262.
- MERTON, ROBERT K. 1938 Social Structure and Anomie. *American Sociological Review* 3:672-682. → A revision of this article appears on pages 161-194 in the 1957 edition of Merton 1949 as "Continuities in the Theory of Social Structure and Anomie."
- MERTON, ROBERT K. (1949) 1957 *Social Theory and Social Structure*. Rev. & enl. ed. Glencoe, Ill.: Free Press.
- MILLS, C. WRIGHT 1943 The Professional Ideology of Social Pathologists. *American Journal of Sociology* 49:165-180.
- OCBURN, WILLIAM F. (1922) 1950 *Social Change: With Respect to Culture and Original Nature*. New ed., with supplementary chapter. New York: Viking.
- PARSONS, TALCOTT 1951 *The Social System*. Glencoe, Ill.: Free Press. → See especially Chapter 7, "Deviant Behavior and Mechanisms of Social Control."
- SMITH, THOMAS LYNN et al. 1955 *Social Problems*. New York: Crowell. → See especially Chapter 1.
- SUTHERLAND, EDWIN H. 1945 Social Pathology. *American Journal of Sociology* 50:429-435.
- THOMAS, W. I. (1909) 1920 *Source Book for Social Origins: Ethnological Materials, Psychological Standpoint, Classified and Annotated Bibliographies for the Interpretation of Savage Society*. Boston: Badger. → See especially pages 16-18.
- THOMAS, W. I. (1917-1937) 1951 *Social Behavior and Personality: Contributions of W. I. Thomas to Theory and Social Research*. Edited by Edmund H. Volkart. New York: Social Science Research Council.
- THOMAS, W. I.; and ZNANIECKI, FLORIAN (1918-1920) 1958 *The Polish Peasant in Europe and America*. 2d ed. 2 vols. New York: Dover. → An unabridged edition of the original five-volume work.
- WALLER, WILLARD W. 1936 Social Problems and the Mores. *American Sociological Review* 1:922-933.
- WARD, LESTER F. 1906 *Applied Sociology: A Treatise on the Conscious Improvement of Society by Society*. Boston and New York: Ginn.
- WHYTE, WILLIAM FOOTE (1943) 1961 *Street Corner Society: The Social Structure of an Italian Slum*. 2d ed., enl. Univ. of Chicago Press.
- WILLIAMS, ROBIN M. JR. (1951) 1960 *American Society: A Sociological Interpretation*. 2d ed. New York: Knopf. → See especially Chapter 10, "Institutional Variation and the Evasion of Normative Patterns."

## SOCIAL PROCESSES

See ACCULTURATION; ASSIMILATION; COHESION, SOCIAL; CONFLICT; COOPERATION; INTEGRATION, article on SOCIAL INTEGRATION; INTERACTION; SOCIALIZATION.



processes are involved in the development of man as a social animal are put into bold relief by such accounts as the *Wild Boy of Aveyron* by Itard (1801). This and similar accounts of so-called feral children suggest that without human contact the infant human organism will never develop his full biological and psychological capacities. Clearly, important events occur in the early, primitive contact that the infant human has with his parents. The nature of these events, the socialization process, has become one of the central focuses of social psychology. If we conceive of socialization in its broadest sense as a process whereby the individual learns the beliefs and values of his social group and learns to adjust his behavior so that it meets the expectations of others in the group, then social psychology can be regarded as the study of socialization and its products. [See SOCIALIZATION.]

Some historical background. There are several traditions that represent different approaches to the study of social behavior, which we can touch on only briefly. The sociological tradition developed by such men as Tarde, Le Bon, Ross, Cooley, and Mead has been concerned primarily with processes mediated by face-to-face interaction. In one way or another this tradition has relied upon one or more of the triumvirate of concepts: suggestion, imitation, and sympathy [see SUGGESTION; IMITATION; SYMPATHY AND EMPATHY; and the biographies of COOLEY; LE BON; MEAD; ROSS; TARDE]. Another important contributor from the sociological side was Simmel, who attempted to distill the forms of social relationships (e.g., superordination and subordination) from the widely different settings in which they appear [see SIMMEL]. A figure who stands alone in importance in the early development of social psychology is the French sociologist Durkheim [see DURKHEIM]. He conceived of the interaction situation as generating emergent norms, or "collective representations," as he called them [see NORMS]. These norms, in turn, act with "exteriority and constraint" in controlling subsequent behavior. This intellectual tradition in sociology has melded into the approach of many contemporary social psychologists.

On the psychological side there have been three major influences: psychoanalysis, behaviorism, and gestalt theory. The psychoanalytic approach conceives of the socialization process as curbing and redirecting the infant's natural appetites, thereby producing an adult with a complex unconscious life, as well as with a conscience that operates to constrain subsequent behavior. Identification, which occurs both in the very early months of life and

in the resolution of the Oedipus conflict, represents, in psychoanalytic theory, one of the principal mechanisms of socialization. The behaviorist approach attributes social learning to the establishment of stimulus-response connections that are mediated primarily by drive reduction; social behavior is seen as a set of ingrained habitual responses and response patterns. The gestalt school, with its emphasis on the organization of the perceptual world of the individual, has focused its study on the meaning of a percept, as well as on the cognitive, conscious thought processes that are involved in interaction. The modern representatives of this school of social psychology, such as Festinger (1957) and Heider (1958), have stressed the tendency for a person to come to see his world and his own behavior and beliefs in a consistent way. [See GESTALT THEORY; IDENTITY, PSYCHOSOCIAL; PSYCHOANALYSIS; THINKING, article on COGNITIVE ORGANIZATION AND PROCESSES.]

In the past these approaches produced a great deal of sectionalism, but much blending is visible today. Workers in the field have come to realize that the various approaches need not be mutually exclusive. Social psychology is coming to be viewed as a general approach to the study of human behavior that rests, on the one hand, on certain basic facts of perception, cognition, motivation, and learning and, on the other hand, on certain basic facts that have to do with the social context which sets boundary conditions for the behavior in the form of the expectations of others. Accordingly, social psychology is in a position to borrow both from traditional psychology and from traditional sociology.

### The general approach and methodology

Any approach to behavior can be cast in the S-O-R paradigm. A stimulus (S) impinges on the organism (O) and produces behavior, a response (R). The ways in which various approaches differ is in how the S, the O, and the R are characterized. Contemporary social psychology differs from traditional psychology in this characterization. Within traditional behaviorism, for example, the characterization of the stimulus and the response have tended to be made in physicalistic terms. For example, in studies of learning, a stimulus is described as a buzzer with a particular frequency or a light with a particular wave length and the response elicited might be a bar-press or the raising of a paw. In general, social psychologists have implicitly assumed at least a two-stage process mediating responses to stimuli. The first is a per-

ceptual stage in which meaning is attributed to some external stimulus configuration and a second stage in which the meaningful stimulus generates subsequent behavior. The emphasis is on the meaning to the person of both the stimulus situation and the response.

Social psychologists have worked in two distinct settings, the everyday world and the laboratory, and have developed a research armamentarium appropriate to both of these settings. By the everyday world we mean the study of social behavior *in vivo*, in some existing social milieu. Social psychologists have studied behavior in factory settings, in military organizations, in housing communities, in delinquent gangs, and so on. Of course, it is impossible to study every aspect of the behavior of the individuals who are enmeshed in such complex situations as these. The social psychologist, therefore, narrows his sights to include only a limited number of aspects of any given everyday-life situation. He may focus on behavior in response to certain aspects of the total stimulus situation, such as the power structure of the group or the communication network that connects the individual member to others in the group. The researcher-observer is typically not in a position to control or alter features of these complex situations and is therefore limited in what he is able to infer with confidence from his observations. If a certain kind of behavior seems to be associated with a particular aspect of the total situation pinpointed by the researcher, it is impossible for him to know, without adequate control over the total situation, whether the behavior was indeed produced by this aspect of the situation or by some other aspect which the researcher failed to take into account. Furthermore, it may be that the behavior itself subsequently served in some way to generate the stimulus conditions; causal direction is difficult to infer. It may also be the case that some unknown set of stimulus conditions is generating *both* the observed set of stimulus conditions and the observed behavior—that is, the observed relationship between the stimulus and the response may be spurious. These are some of the methodological shortcomings to which research in field situations is heir. [See EXPERIMENTAL DESIGN.]

Much of the field research in social psychology has been devoted to the study of such specific problems as the way in which supervisory practices affect morale in a particular organization or how certain propaganda appeals affect the vulnerability of a particular audience. Field methods are eminently suited to finding answers to specific, press-

ing problems of this kind. The laboratory serves—or should serve—a different purpose. Because of his ability to manipulate, control, and purify the stimulus conditions confronting the person who is serving as a subject in the experiment, the experimenter acquires the ability to study behavior in more general terms. His ability to infer causal connections between the stimulus conditions that he creates and the behavior of his experimental subjects helps in developing theories regarding the underlying psychological processes—the nature of the *O*. He attempts to do this both by establishing specific stimulus conditions confronting his experimental subject and by limiting the response alternatives available to the subject. His inferences are generally made by comparing the effects on behavior of variations in the stimulus conditions with the same set of response alternatives available to the subject.

The fact that the laboratory is often used to create a minuscule social world has led some people to assume mistakenly that the main purpose—in fact, the sole purpose—of the laboratory is to try things out on a small scale before applying them to a larger, real-world context or to understand the large complex situation by re-creating a smaller situation that has the same order of complexity as the larger one, a rather ambitious program to say the least. The laboratory research worker who is investigating basic social psychological processes is not concerned with making empirical generalizations relating a specific set of stimulus conditions to a specific response, but rather with establishing conceptual laws about behavior. The mistaken testing-ground view of the laboratory is a source of much confusion. For example, it leads to the expectation that the laboratory worker should provide people in policy-making positions with immediately useful insights and specific empirical predictions concerning some specific aspect of their particular world. The field setting and the laboratory are geared to answering quite different questions. The field setting is much richer in content, but this richness also makes for tenuous inferences.

The technology used in any field study or laboratory experiment will be dictated by the specific problem at hand. Typically, a field study involves a combination of careful observation of behavior with some form of verbal report, questionnaire, or interview responses. Whereas social psychologists who engage in field studies devote a great deal of effort to refining these response measures, the laboratory worker devotes the greater part of his efforts



to refining the stimulus conditions. This difference in emphasis is due to the distinct differences between the two settings. In the field study, the investigator studies responses to existing stimulus conditions over which he has no control, whereas in the laboratory he sets out to create the stimulus conditions and usually permits only a narrow latitude of possible responses. As we review some of the research we shall see examples of individual items of technology. [See *INTERVIEWING, article on SOCIAL RESEARCH*; *OBSERVATION, article on SOCIAL OBSERVATION AND SOCIAL CASE STUDIES*.]

**The action sequence.** A general approach to behavior that incorporates aspects of a number of other approaches is that of the action sequence. Any action engaged in by the person is designed to come to terms with his changing stimulus environment so that his outcomes (satisfactions) will be maximized. He will want to transact with objects having positive value for him and will avoid objects having negative value. As compared with lower organisms, the stimulus vista of the human being is extremely broad, consisting of a wide variety of potentially satisfying and potentially punishing objects and events. Much of this complex set of value-object relationships, or attitudes, is acquired through the socialization process.

Any given act involves a decision, however simple and quickly it is made, between alternatives, each having consequences for the person. His decision—even if it is between acting or maintaining the status quo—will be predicated upon how the person sizes up the outcome potential of each alternative. A postdecisional accommodation process occurs after he commits himself to one of the action alternatives. Behavioral economy and the exigencies involved in dealing with the world around him require that the person make relatively quick decisions and then follow through. He learns to live with his decisions, since undoing them is often costly, if not impossible. Also, a decision implies giving up the positive outcomes associated with the nonchosen alternative and suffering the negative outcomes, if any, associated with the chosen alternative. A decision is forced upon the person, since he is unable to simultaneously enjoy the values associated with the various alternatives that are open to him. Vacillation is not an efficient solution. Postdecisional accommodation has been studied extensively by Festinger (1957) and others. The person's predecision stance toward information about the alternatives is an open or vigilant one, whereas his postdecisional stance is selective and defensive. This basic antinomy between vigilance and defensiveness is a fundamental characteristic

of behavior. We shall review the general field of social psychology within this action-sequence framework which considers both the predecisional and postdecisional phases of the act.

### Perceptual processes

Social psychologists have tackled the age-old and difficult problem of perception. The problem stated simply is: How is the welter of stimulus information that impinges on the person at every moment organized by him and made meaningful? The act of perception is a decision to classify some stimulus event in some particular category. This response will be determined by the total behavior context in which the perceptual act occurs. A configuration conveying stimulus information of a round red object with a stem sticking out of a small depression on the surface may in one context be perceived primarily as something to eat and in another context as something to throw. The person's "definition" of the situation will be determined, in part, by prior social learning. This situational definition consists of a set of expectations as to what stimulus objects are likely to be present within the general behavioral context and the values that the person holds as he contemplates transaction with the stimulus. A percept is a decision and as such can be analyzed within the action-sequence framework. In Bruner's terms (1957) a percept is a "solution" to a problem, the constituent parts of which are the cues provided by the object and the expectations and desires provided by the person. A good solution is defined pragmatically as one in which subsequent transaction with the object confirms the original expectation and, hence, provides satisfaction for the original desire. The perceptual process serves not only to facilitate transaction with positively-valued objects but also to prepare the person to avoid negatively-valued ones. He is continually alert to changes in his stimulus environment that may signal positively-valued or negatively-valued consequences for him.

Most of the research has been in the form of demonstrating the interaction between the stimulus and the person's values in one or another behavior context. There has been little work, however, on the exact nature of the interaction that would help to define the limiting conditions under which the interaction will take place and those conditions under which the interaction will be maximum. The findings that support this view of the perceptual process are reviewed in detail elsewhere [see *PERCEPTION, articles on PERSON PERCEPTION and SOCIAL PERCEPTION*].

A particular relationship deserves special mention here since it highlights the functional, decisional nature of perception. Under conditions of impoverished stimulation, the subject's visual threshold for positively-valued stimuli is lower than it is for neutral stimuli—that is, less stimulus information is required by the subject to recognize positively-valued objects. The findings for threatening, negatively-valued stimuli show a somewhat different pattern. Negatively-valued stimuli that can be avoided show a lower threshold than neutral stimuli, whereas negatively-valued stimuli that cannot be avoided show a raised threshold. Both of these tendencies can be interpreted as attempts at avoidance. Where the subject can take action, he is vigilant toward the threatening event, whereas where he can do nothing to avoid it he is ostrichlike in this tendency to deny its existence. [See ATTENTION; PERCEPTION, article on PERCEPTUAL DEPRIVATION.]

**Person perception.** The study of person perception is, of course, of special interest to social psychologists. The way in which one person's acts are interpreted by another will determine the course of their future interaction—that is, whether or not mutual dependence and mutual influence develop. The general findings in this area are consistent with our functionalist view of perception. The total stimulus-and-value context appears to induce expectations as to the qualities or traits possessed by the stimulus person as well as to induce attributed intentions. The fact that a person is known to possess a particular characteristic induces the expectation that he will possess other characteristics that are usually related to the one that he is known to possess. The inference made about the stimulus person from any given act is made in terms of the context of the act. Was the act justified in the eyes of the perceiver—that is, was the act appropriate to the stimulus situation attributed by the perceiver to the stimulus person? For example, if the stimulus person behaved aggressively within a context that appeared to the perceiver as one in which aggression is a perfectly appropriate response, the stimulus person might not be perceived as aggressive. He would tend to be perceived as aggressive only if another, nonaggressive response seemed more appropriate to the stimulus context. On the other hand, if the context is one in which an aggressive response seems particularly appropriate and the person responded in a nonaggressive way, the perceiver might attribute meekness to the stimulus person. The general principle operating here is that the intention attributed to the stimulus person will be inferred from the stimulus person's

behavior judged in terms of the behavioral alternatives open to the stimulus person (as seen by the perceiver). Various aspects of the total situation will determine the judgment by the perceiver as to the response alternatives open to the stimulus person—for example, the status of the stimulus person vis-à-vis the person toward whom he is acting, the sequence of interactions immediately preceding the segment of behavior in question, and idiosyncratic qualities of the perceiver himself.

**Interaction.** Man's social nature is presumably molded through interaction. Since interaction is between persons as each is perceived by the other, knowledge of the process of person perception is an important and basic key. We form an impression of a person from his interaction with us or with others and our impressions in turn condition our interaction with him. The research on social influence and other group processes, which we shall review below, starts either by assuming that the actors have certain impressions of each other or by experimentally inducing the impression, *de novo*, as an independent variable. The student of person perception examines the process in reverse by treating interaction as the independent, experimental variable and examining aspects of the interaction situation that produce various impressions. Both of these approaches complement each other, and knowledge gained from each will add to our understanding of the nature and substance of social life.

### Social comparison processes

Other persons can act as sources of information or as direct mediators of reward. The child's parents provide him with food, warmth, and shelter and also with certain knowledge he requires to engage in the business of living within his social milieu. All problems in social psychology reduce to studying implications of either one or the other kind of dependence for information and for directly mediated effects (outcomes, rewards, satisfaction).

A given act involves a beginning state, a desired end state, and a path or sequence of behaviors designed to move the individual from start to finish. Any act, however trivial, can be analyzed within this framework. Kurt Lewin referred to the possible action sequences existing for the person at any given moment as the person's "life space." There may be a considerable amount of uncertainty attached to various aspects of an action sequence. The person may be uncertain as to whether his present level of satisfaction is reasonable or whether the end state of the sequence will indeed furnish him with a higher level of satisfaction. He



also may be uncertain as to whether or not he has the ability to act in the manner required to realize the end state or to transact with the end state in order to derive the satisfaction he anticipates. To the extent that the situation is novel, his uncertainty will be high. The person's world is in constant and irregular motion, usually taking unpredicted turns. In trying to reconcile these changes with his own desires, he is continually confronted with decisions and tries to choose the action alternative that he hopes will yield him the most pleasure or the least pain. To the extent that he is uncertain about what to do or how to do it he will be oriented toward getting information to reduce his uncertainty.

The person may be uncertain about his present outcome level—that is, whether or not he is receiving adequate satisfaction in his present situation. What kind of measuring stick does he use to determine how satisfied he is? There appear to be three bases upon which he may “measure” his satisfaction: his own past experience, the various alternative outcome states potentially available to him, and the experiences of other people in the same or similar situations. It is this last basis that is of particular interest to us here. Where comparison persons are available as points of reference against which he can compare aspects of his own situation, he will make use of this information to the extent that more reliable information is lacking.

**Reference persons.** There are three basic kinds of information other persons can provide: advice, impressions, and knowledge as to relative standing. Advice is any information given to another through “instruction.” Impressions are given about the person by another through a process we shall call “reflected appraisal.” Standing refers to the person's relative position in a group, with regard to some attribute, and the process whereby he determines his standing we shall call “comparative appraisal.” Most of the research on social influence has been concerned with the last-mentioned process, comparative appraisal.

Two general types of reference persons can be distinguished: the “expert” and the “co-oriented peer.” An expert is someone who, by virtue of his experience, has special knowledge as to how the person might effect a favorable change in outcome level. A co-oriented peer is someone who shares the life situation of the person, someone who has a similar outlook and value orientation. In measuring his present or future satisfaction level the person tends to refer to others like himself. Merton and Kitt (1950) demonstrate this rather well with data on the degree to which certain groups of en-

listed men felt deprived during World War II. They found, for example, that married soldiers felt more deprived as a result of being in the army than did single soldiers. This, they discovered, was attributable to the fact that married soldiers, in measuring their outcome level, compared themselves with other married men, whereas single soldiers compared themselves with other single men. Since all able-bodied single men were in the army, the single soldier did not feel deprived relative to his co-oriented peer group. Married soldiers, on the other hand, knew that a large proportion of their peer group had evaded the draft and were enjoying the comforts of home life. They therefore felt deprived. The co-oriented peer helps the person to measure his present or potentially available outcome states, whereas the expert is a source of information on how to move from one state to another. [See REFERENCE GROUPS.]

**Social influence.** The research on social influence has been conducted in three very different laboratory settings, and our review will examine each of these in turn, within the general framework outlined above. The first setting is one in which the subject finds himself in a face-to-face confrontation with other people with whom there is some sort of a discrepancy. The second setting is one in which the subject is merely informed that he is in disagreement with others and has little or no opportunity to communicate with them. The third setting is one in which a single communicator attempts to persuade an audience.

### Communication and consensus in groups

The work of Allport (1920) brought the study of group effects on individual judgments into the laboratory. Allport's work was preceded by a tradition dating back to the turn of the century, in which group effects on individual performance were demonstrated. Allport found that a person's judgment about an ambiguous stimulus can be influenced by the judgments of other people about that stimulus. In subsequent work, Sherif (1935) found that this modified judgment, the norm, tends to persist even in the absence of these other people. We can see the close connection here with Durkheim's concept of collective representation. [See GROUPS, article on GROUP FORMATION.]

The advent of World War II spurred the development of social psychology, especially in the area of social-influence processes. The next important milestone was provided by the work of Kurt Lewin and his students. Lewin realized very early the importance of systematic theory in studying behavior, and he also had an abiding faith that even

complex social behavior could be studied systematically in the laboratory. Of particular concern to us here are his studies of group decisions. These studies, which were undertaken during World War II, suggested that when an individual reaches a decision in a group situation, the decision appears to have a compelling quality in sustaining certain consequences of that decision (Lewin 1947). Some doubt has been cast on the conclusions reached in these early experiments, by some more recent work of Bennett (1955). In her attempt to replicate with more careful controls the earlier group-decision experiments, Bennett found that it was the decision itself that carried the day rather than the fact of having reached a decision in a group. Bennett discovered, however, that the individual's attitude toward the decision was positive to the extent that there was consensus in his group. This, then, may have been the crucial mediating factor in Lewin's experiments—namely, that where a group decision was made, the individual group member was made aware of the group consensus. [See CONFORMITY; CONSENSUS; GROUPS.]

**Attraction of the group.** Lewin's "discovery" of the power of the group inspired the work of a number of his students. One set of studies that was so inspired was undertaken by a group working with Lewin that was led by Festinger. The first of these was a field study by Festinger, Schachter, and Back (1950) of social pressures in a student housing community in the Boston area. Within the community, "natural" groups were created by the architectural design of the community; families were thrown together because of the way in which the individual apartments were arranged. The central finding of the study grew out of the fortuitous occurrence of an important, controversial issue that arose while the study was in progress. The investigators found that the degree of unanimity of opinion on the issue within each of the architecturally circumscribed groups was a direct function of the over-all attraction of the individual members to these groups. There is a strong suggestion here that the more attractive a group is for an individual member, the greater will be the tendency for the member to be influenced by the opinions of others in the group. This interpretation led to more general speculation about the nature of social influence processes and eventuated in a program of laboratory research that occupied Festinger and his students for a number of years. Typically, the experiments used college-student subjects engaged in a discussion during which their behavior was carefully observed and meas-

ured. Three types of behavior were recorded: an attempt by the subject to influence someone else in the group, an attempt by the subject to reject someone in the group, and a change of opinion by the subject on the issue that was being discussed. The amount of each of these three kinds of behavior, it was reasoned, would indicate the strength of the tendency for the group to reach a consensus. With these behaviors as the dependent variables, various aspects of the group situation were manipulated experimentally. The suspected relationship found in the housing study mentioned above—namely, that the greater the attraction of a group for its individual members, the greater will be the tendency for members of the group to reach a consensus—was verified (e.g., Schachter 1951; Festinger et al. 1952). It was also found (e.g., Schachter 1951) that the greater the relevance of the issue to the purpose of the group, the greater will be the attempts to reach consensus. To the extent that the members of the group are able to refer their opinions outside the group for validation, there will be a weakened tendency to achieve consensus in the group (Festinger & Thibaut 1951). Studies by Festinger and Thibaut (1951) and Gerard (1953) found that the greater the discrepancy in opinion between members of a group, the greater will be the tendency for them to influence each other. To the extent that a given group member is close to the modal opinion in the group, he will attempt to influence those who are deviates to change their opinion, so as to achieve greater group consensus (Festinger et al. 1952; Gerard 1953). Reaching consensus in a group involves getting deviates to change their opinions. When such attempts fail, consensus can be achieved by rejecting these deviates. Any basis that may exist for conceiving of the deviate as not being a relevant reference person will tend to enhance rejection as a method for achieving consensus (Festinger & Thibaut 1951; Gerard 1953). There is also evidence that the individual will be more vulnerable to influence from someone who is an expert than from someone whose ability is equal to or less than his own (Festinger et al. 1952).

All of these experiments growing out of the housing study were designed to study information dependence that is engendered when the individual is uncertain about some belief or opinion that he holds. Other people, it was assumed, are used merely as reference points against which the person may check his opinion. The process based on this kind of referral, comparative appraisal, can occur with or without actual face-to-face confron-



tation. When, however, people do confront each other face-to-face, reflected appraisal may also take place. Here the person is responding to what the other person may think of him for being a deviate; he will respond not in order to reduce uncertainty, but in order to maintain or improve his status in the other person's eyes. In these experiments the subjects were confronting each other face-to-face, and it is likely that their responses were in part due to reflected appraisal. It is not clear, for example, why attraction to others in the group should increase the person's tendency to refer his opinions to that group, unless we assume that reflected appraisal is operating—that is, that individual members are concerned with maintaining their status in the group. Where the person is concerned with reducing the uncertainty with which he holds a belief, referral should be based upon how much expertise the person attributes to the reference person and not to how attractive he is. There is another possibility that may account for the relationship between comparison and attraction. Insofar as the aspect or judgment being compared with others in the group has anything to do with present or future satisfactions—that is, outcome perspective—there will be a greater tendency to rely upon the judgments of others to the extent that they are perceived as co-oriented in this respect. It may be that where a person is highly attracted to others in the group, he assumes that they are co-oriented peers. Another possibility that was addressed experimentally by Berkowitz (1957) is that an individual will attribute expertise to someone he likes. In any event the original problem with which these studies began—namely, the relationship between the tendency to reach a consensus in a group and attraction by the members to the group must be mediated by either attributed co-orientation or attributed expertise. A great deal of work remains to be done in order to understand the nature of the mediational processes involved. [See COHESION, SOCIAL.]

**Further studies of social comparison.** An important field study conducted by Newcomb (1943) in the middle and late 1930s at Bennington College, a women's college, offers striking evidence regarding the attainment of consensus in an existing group. Newcomb traced the changes in attitudes toward political affairs that occurred among Bennington students during their four years of college. The typical Bennington freshman in 1935, the year the study began, came from a wealthy northeastern conservative home and reflected this background in her attitudes upon entering Bennington. Most of the faculty were left of center

in their political attitudes. As the situation developed, it was "in" to be a liberal, and most of the upperclassmen expressed liberal attitudes. The longer the student remained at Bennington, the more liberal her attitudes became. That this was not a mere overt public avowal, but a reflection of her actual attitudes, is attested to by a follow-up study of the attitudes held by these Bennington students thirty years later (Newcomb 1963), in which it was found that there was very little reversion to the conservative attitudes held by these women when they were entering Bennington freshmen nearly three decades earlier.

A series of experiments initiated by Schachter (1959) studied comparative appraisal as mediated by attributed co-orientation. Whereas the earlier experiments were concerned primarily with the comparison of beliefs, Schachter's experiments were concerned with comparison of emotions. Emotions are strong attitudes and, as such, have a belief and a value component (see the discussion below on attitudes). The value component has associated with it a substrate of physiological arousal. Since the value component is paramount in the case of an emotion, a person who is uncertain about some aspect of an emotional experience he is having will seek out co-oriented peers—that is, others undergoing or about to undergo the same emotional experience. Schachter found evidence for this in an experiment in which he threatened subjects with a strong electric shock. In subsequent experiments additional evidence was found that emotional uncertainty does lead to information dependence on co-oriented peers.

There is also evidence to indicate that the performances of other people are sometimes used by the person to estimate the quality of his own performance. A study by Chapman and Volkmann (1939) demonstrated a "group effect" on the person's aspiration. Where the subject had no clear expectation as to what his performance would be on an unfamiliar task, he utilized information about the scores others had made on the task, in setting his own level of aspiration. On the other hand, where the subject had had considerable experience with a particular type of task and could therefore estimate fairly accurately how he would perform on a similar task, knowledge of the scores others had made had little or no effect on his performance aspiration.

Other studies have shown that when a task is a competitive one, the person will seek his own level in choosing a competitor. This finding is taken by Festinger (1954) as evidence for a tendency on the part of the person to compare his

performance to the performance of others as a way of evaluating his ability.

### The conformity conflict

The research on consensus attainment in groups was, for the most part, concerned with events taking place in the group as a whole and did not pay specific attention to the individual member's reactions to a discrepancy confrontation. And elegantly simple experimental situation designed by Asch (1956) provides a technique for bringing the individual's reaction into sharper focus. This is accomplished by confronting the subject with unanimous disagreement from a group of peers (who are actually accomplices of the experimenter) on a very simple, unambiguous visual judgment. The judgment involves matching the length of a single line to one of a number of comparison lines. The subject confronts repeated disagreement from "the others" on a series of trials. Asch found that there was a tendency for some subjects to yield to the group judgment.

In a situation like this the subject assumes that the others possess approximately the same perceptual capacities as he does and are able and willing to report accurately what they see. How, then, can there be a discrepancy like this? These considerations make up the informational side of the conflict. There is also the problem of standing out like a sore thumb, the subject's concern with the kind of impression he is making on the others if he assumes that they expect him to agree with them. If he disagrees with them he may lose status in their eyes. We can anticipate, then, that both comparative and reflected appraisal will be operating in the situation. An attempt to separate these two processes was made by Deutsch and Gerard (1955), by comparing a treatment in which the subject made his judgments in full view of the experimental accomplices with a treatment in which the subject made his judgments within the privacy of an isolation booth. Less yielding was found when the subjects were isolated from each other than when they were face-to-face, suggesting that reflected appraisal is a strong component process of the face-to-face situation. A number of other experiments have substantiated these results. [See CONFORMITY; CONSENSUS.]

Information dependence on others appears to increase with increasing ambiguity of the judgmental material (e.g., Asch 1956). If we assume that an ambiguous stimulus will lead to uncertainty, this finding supports the basic assumption underlying comparative appraisal: uncertainty leads to information dependence. When there is

information dependence a clear "expert effect" has also been found—namely, that when confronting an ambiguous situation, the subject is dependent upon the others to the extent that he perceives them to be expert in the kind of judgment he made (Samelson 1957). One might be led to assume by the results of these experiments on relative ability that the person with high ability has little difficulty in remaining steadfast, that he simply discounts the discrepant information as being unreliable, and that it is the low-ability person who is in trouble. An attempt to explore the psychological impact that disagreement has on individuals of different ability was made (Gerard 1961), using physiological responses to measure the impact of disagreement. The findings in these studies indicate that the greater the person's ability, the greater will be the impact of disagreement with a group of peers. This result may be interpreted by assuming that the higher the person's ability, the greater the conflict between the two sources of information available to him, the information from his own senses and the group consensus. The greater his ability, the more credibility will he attach to his own judgments which are in disagreement with another highly credible source of information, the group consensus.

In an experiment involving relatively strong attitudes, Kelley and Woodruff (1956) found that disagreement with a group of co-oriented peers was more effective in changing the subject's attitude than disagreement with a group that was represented as not being co-oriented with the subject.

In a number of investigations, explicit effect dependence was induced by making the subjects dependent upon one another for certain outcomes (Deutsch & Gerard 1955). This effect dependence produced greater conformity to a false consensus, an outcome presumably mediated by the subject's desire for acceptance in situations where the others are potential mediators of satisfaction for him. A particularly crucial finding that is relevant here comes from a study by Kelley and Shapiro (1954), in which a negative correlation was found between conformity and the acceptance of the subject as a group member. Evidently, the less the subject's acceptance, the more anxious was he to please the other members of the group. Additional supporting evidence was found by Harvey and Consalvi (1960), in whose study the most and least accepted members of a group conformed less than a member who was only partially accepted. We can interpret these results by assuming that the person of highest status does not conform because



he is assured of his status, whereas the person of lowest status does not conform because he sees no possibility of improving his status.

Other studies have been concerned with the role of commitment in the conformity conflict. Deutsch and Gerard (1955) discovered that when the subject committed himself to his judgment before being apprised of the judgments of the others in the group, he yielded much less to the group consensus.

In terms of the action sequence, we would expect that the decision to yield to the group or remain steadfast would have postdecisional consequences. Evidence reported by Gerard (1965) suggests that postdecisional accommodation does indeed take place. A conformer will accommodate both by increasing the degree to which the group is attractive to him and by attributing expertise to them, whereas a deviate will tend to reduce the perceived attractiveness of the group and attribute lower expertise to them as a way of justifying his own behavior.

### Attitude change

We will attempt a brief but broad overview of the area of attitude change in order to show how our general framework applies to this area as well. We shall examine, in turn, characteristics of the three aspects of the persuasion situation: the communicator, his message, and his audience. This is the framework used by Aristotle in his *Rhetoric* and by Hovland, Janis, and Kelley (1953) in their classic work. [See ATTITUDES, article on ATTITUDE CHANGE.]

An attitude designates the outcome potential the person associates with some object or class of objects. He consequently has a tendency either to approach or avoid that class of objects. The structure of an attitude can be represented as a syllogism in which the minor premise is a belief, the major premise is a value, and the conclusion is the attitude itself. Thus, a prejudiced person might be characterized by the following syllogism as part of the structure underlying his attitude:

Negroes are lazy.

Laziness is bad.

Therefore, Negroes are bad.

An analysis of this syllogism shows that "Negroes" is the minor term, "laziness" is the middle term, and "bad" is the major term. Many attitudes, such as anti-Negro prejudice, have complex structures that include sets of parallel syllogisms all having the same conclusion and a "deep" vertical structure with interlocking premises—that is, working backward we may find that the premises in a first-

order syllogism may be conclusions of one or more second-order syllogisms, and so on. There has as yet been no attempt to describe systematically the structure for a given person, although case studies such as those reported in Smith, Bruner, and White (1956) represent a start in that direction.

**The communicator.** Propaganda appeals are directed toward changing either the belief or value premises underlying an attitude. The credibility attached to a message will depend upon the degree of expertise or co-orientation attributed to the communicator, expertise being more important where a belief is a target of the message and co-orientation being more important when a value is the target. Experiments (e.g., Aronson et al. 1963) have demonstrated a positive relationship between the credibility of a message and the amount of expertise attributed to the communicator. An experiment by Weiss (1957) provides indirect evidence for the relationship between credibility and co-orientation of the communicator.

An aspect of the situation that is closely related to co-orientation has to do with the intent attributed to the speaker by his audience. If he is perceived as wanting to influence them, it may be in pursuit of some value he holds that his audience perceives as inimical to their own values. Therefore, to the extent that the communicator is seen as not attempting to influence his audience, the effect of any attributed lack of co-orientation would be minimized. An experiment by Walster and Festinger (1962) compared the effectiveness of a communication when it was accidentally overheard with a situation where the communicator was aware of his audience (and his audience of him) and found that there was greater attitude change in the direction of the message when it was merely overheard. [See COMMUNICATION, MASS.]

**The communication.** An aspect of the message that has been given some attention concerns the question of whether it is more effective for the communicator to present only his side of the argument or to present the other side of the issue as well. The results of this research are inconclusive. There does seem to be some evidence, however, that when a person changes his attitude in response to a two-sided communication, this change will be more likely to sustain itself than if he changes in response to a one-sided communication. It may be that a two-sided communication presents the person with a definite decision, and the communication's ability to sustain a change may be due to postdecisional accommodation. [See PERSUASION.]

**Primacy versus recency.** When presenting both sides of an argument, is it more effective for the

communicator to present his side first or second? This question reduces to whether a "primacy" or a "recency" effect will dominate. Again, the research findings do not offer unequivocal support for the predominance of either effect. Some studies have shown a primacy effect, whereas others have shown a recency effect. When the subject commits himself to an opinion after hearing one side of the issue before hearing the other side, a strong primacy effect occurs (Hovland, Campbell, & Brock 1957). This effect can be interpreted as being caused by postdecisional accommodation to the "premature" opinion commitment. A study by Miller and Campbell (1959) suggests that some of the contradictory evidence has been due, in part, to methodological problems that arise in presenting the messages and measuring their effects. As with a number of other problems we have discussed, a great deal more research remains to be done.

*Implicit versus explicit conclusions.* Another question for which there is no satisfactory answer is whether or not the communicator should draw his conclusion explicitly at the end of his talk or whether he should let his audience participate by "putting two and two together" for themselves. Aristotle suggests that a message that allows the audience to draw its own conclusion will be more effective. In our terms, conclusion drawing is a decision which, because of postdecisional accommodation, will sustain the effect of the message—if the audience draws the intended conclusion. A study by Hovland and Mandell (1952), however, suggests that conclusion drawing by the communicator is the more effective technique. Cooper and Dinerman (1951), on the other hand, find that conclusion drawing by the communicator blunts the effectiveness of a message. It is probable that particular circumstances will determine which technique is more effective, and the job of future research is to learn what these circumstances are.

*Conditioning effects.* Staats and Staats (1958) used a classical conditioning situation to manipulate the belief premise of an attitude and then measured the degree to which the middle term transferred value to the minor term that had originally been associated only with the major term. The results show that mere contiguity with positive or negative values will impart a positive or negative value to a term that was neutral to start with. Razran (1940) conditioned subjects to respond positively or negatively to formerly neutral stimuli, with a subtle technique of associating the neutral term with faint, but perceptible, unpleasant odors or the circumstance of munching away on

an appetizing free lunch. Powerful effects such as these are probably being induced without our awareness during the course of a normal day. These effects, which appear to occur without our awareness, are paralleled by effects that are based on conscious, cognitive inference. Where beliefs are shown by the communicator to further certain values, attitude change can be produced (e.g., Di Vesta & Merwin 1960).

*Boomerang effects.* There is evidence that messages that embody very strong emotional appeals may boomerang (e.g., Leventhal & Kafes 1963). The reasons for this are not very clear, although there is some suggestion in the data that strong value appeals induce defensiveness, especially where the attitude is difficult to change.

*The audience.* Obviously, the effectiveness of any message will be determined by the degree to which it is successfully tailored to its intended audience. A communication intended to induce new toothbrushing habits in an audience of eight-year-olds would not have the same impact on an adult audience. Factors such as the pre-exposure attitudes of members of the audience, the degree of familiarity with opposing points of view, their education and background, and their degree of commitment to their initial positions will all affect the degree to which they will be susceptible to a given message.

*Pre-existing attitudes and forewarning.* Hovland (1959) points out that one of the features distinguishing field studies from laboratory experiments on attitude change is the lack of forewarning of the audience that exists in the laboratory situation. In the "real world" a person selectively exposes himself to one or another propaganda missile, whereas the laboratory subject is a captive audience. Hovland suggests that this difference may be one of the factors accounting for the greater relative effectiveness of propaganda attempts studied in the laboratory, where the subject can be caught off guard, so to speak. Studies (e.g., Freedman & Sears 1965) have demonstrated that a forewarned audience is a forearmed audience and is less vulnerable to the communicator's message than an audience that has not been forewarned. An ingenious experiment by Festinger and Maccoby (1964) indicates that partially distracting the audience while the message is being presented will tend to increase the audience's vulnerability to the message.

Research by Cooper and Jahoda (1947) suggests that the prior attitude of an audience will affect how a message is interpreted. This is in line with Asch's suggestion (1948) that the context in



which a statement is heard will, in part, determine its meaning for the person. Since the effectiveness of a message hinges on its getting across to the audience, interpretation operates as an important influence in the situation. A study by Hovland, Harvey, and Sherif (1957) demonstrates that one of the factors involved in the interpretation process is the degree of extremity of the message from the subject's own attitude. Messages that are close to the subject's own position tend to be assimilated to that position, whereas messages that are inconsistent with his position are seen as even more discrepant than they actually are.

In a series of studies McGuire (1964) has shown, using an immunological analogy, that small doses of the propaganda prior to an actual full-scale attack will stimulate the person's defenses against the message. McGuire used cultural truisms such as "You should brush your teeth at least twice a day" as the opinions that were attacked. He points out that the average person's prior preparation to defend these truisms is minimal, that they have grown up in a "germ-free" environment and are therefore particularly vulnerable. Whether small doses of propaganda are able to stimulate a person's defenses regarding controversial issues remains to be seen.

*Role playing.* A procedure that has been used extensively both in training and in psychotherapy requires that a person play the role of someone else. This experience presumably gives the trainee or the patient some idea of what it is like to be the person whose role he has taken. He is able to develop an approximate version of the other person's world and in so doing comes to develop something of an understanding of what it is that is producing the other person's attitudes and behavior—that is, he becomes co-oriented with him. The first systematic attempt to investigate the effect of role playing was made by Janis and King (1954). The question they asked was, Is it the fact of verbalizing statements having to do with the other person's point of view or is it simply a matter of being exposed to that point of view that produces the change in attitude claimed for role playing? They were asking, in effect, whether role playing is really any more effective than the typical attitude change procedure that exposes the subject to an opposing viewpoint. Their data suggest that role playing does add something over and above merely being exposed to the message. The decision to play a role, like any other decision, involves all of the elements of the action sequence. The person is confronted with a choice of whether or not to play the role. If he chooses to play the role, he

must somehow come to terms with this decision. The efficacy of role playing may, therefore, hinge on the process of postdecisional accommodation. [See *ROLE*.]

We would expect attitude change to be greater to the extent that there was little other justification for playing the role. The person with little justification would be hard put to justify taking a discrepant viewpoint. One way of justifying the role-playing decision is to accommodate by finding the role itself attractive. This would tend to induce a positive attitude toward the discrepant viewpoint being advocated. Evidence that attitude change is inversely proportional to the incentive given to the person for playing the role is provided by Brehm and Cohen (1962). These experiments used monetary incentives to induce the subject to advocate an opinion that was discrepant from the subject's private opinion and found that the subject changed his private opinion more with a lower incentive. Other experiments have used inducements such as helping the experimenter in his research or the warmth and friendliness of the experimenter and found the same inverse relationship between attitude change and the amount of incentive. The principle operating here is that the greater the incentive, the less will be the necessity for the person to justify his decision. Other studies, using negative incentives—that is, coercion—have discovered that the greater the perceived freedom of choice in playing a role, the greater the amount of postdecisional attitude change. An important experiment by Mills (1958) found that when a person engages in forbidden behavior, the less the incentive for doing so, the more lenient does he become in his attitude toward the forbidden behavior. If the person decides not to engage in the forbidden behavior, the greater the original incentive for engaging in the behavior, the more severe does he become in his attitude toward the forbidden behavior. Festinger and Freedman (1964) suggest that behavior justification of this kind may be an important mechanism in the development of moral values. When a child engages in behavior prescribed for him by his parents or when he decides not to engage in behavior that is forbidden by his parents, to the extent that justification for his decision is low he may subsequently justify the behavior by coming to believe that it is the morally correct way to behave.

### Social interaction

The study of social influence shades over into the more general study of interaction. In the research discussed above, the focus has been limited

to an influence attempt by one person upon another with no influence being exerted in return. Where two individuals confront each other face-to-face, typically the behavior of one is influenced, at least in part, by the behavior of the other in a *seriatim* fashion. Each, to some extent, controls either information or effects (rewards, outcomes) for the other. It is through this control, as we have seen above, that both behavior and attitudes are shaped.

There have been several approaches to the study of the *seriatim* situation. A line of investigation started by Greenspoon (1955) has proved fruitful. The paradigm in these experiments is for the experimenter to reward the subject for saying certain kinds of words by either a nod of his head or saying "Good" or "Mmmmm." This is done in such a way that the subject is usually not aware that he is being rewarded in this way. For example, the subject might be rewarded for saying plural nouns or naming types of fruit. The subject's tendency to say these words after the experimenter no longer rewards him for doing so is shown to increase as a function of reward. Cohen, Greenbaum, and Mansson (1963) have shown that a period of prior social deprivation enhances the effect of this kind of reward. We can conclude from the considerable work carried out in this tradition that mutual control is probably governed in part by the administration of social reinforcements.

Sidowski, Wyckoff, and Tabor (1956) designed a "minimal" social situation in which to study the applicability of learning principles derived primarily from animal studies to the social contingencies that exist in mutual human interaction. The situation utilizes two interdependent subjects, each being unaware of the presence of the other. Each subject makes a choice between two possible responses. One choice rewards and the other choice punishes the other subject. Subjects soon learn, in spite of their being unaware of each other, to choose the mutually rewarding pair of responses. Further studies (e.g., Kelley et al. 1962) have explored factors that enhance or detract from the two subjects hitting on mutually rewarding responses. Informing the subject that his behavior affected the rewards of someone else and that someone else affected his rewards appears to enhance the rapidity with which the mutual contingency is "solved." These experiments seem to have abstracted the minimal elements involved in cooperative behavior, where the person can best serve himself by helping the other. Responses that reward another will tend to be reciprocated and will therefore be rewarded in turn.

Studies of a more complex nature have been carried out in which the joint payoff is arranged so that mutual trust is necessary in order for responses to be mutually rewarding. It is possible in these "games" for each player to "double-cross" the other and, if he is successful, to achieve a reward that is larger than the reward that would result from a mutually cooperative response. Thus, there is a temptation to be asocial. Also, in the process, the other player suffers at the expense of the first player getting a very large reward. A number of investigators (notably Deutsch 1958) have studied conditions under which a social, cooperative response will be made in spite of the temptation that exists to make an asocial response—that is, conditions under which mutual trust will develop. Some factors that appear to foster trust are the possibility for communication, benevolent behavior on the part of the other person, the presence of a common enemy, the absence of mutual threats, and a balance of power.

In a programmatic attempt to study techniques of augmenting one's power in an interaction situation, Jones (1964) has examined conditions under which one person in a relationship will tend to be ingratiating. Ingratiation is illicit behavior in which the person attempts to improve his status in the relationship by appealing to extrarelational values held by the other person. Any relationship is defined by the kinds of rewards each person supplies for the other. To the extent that one of the parties cannot supply the rewards expected of him, the other will tend not to reciprocate. If he is badly in need of the outcomes the other person can potentially provide, he may resort to illegal means in order to obtain them, ingratiation being one possible strategy. The person will attempt to make himself attractive to the other by being a "yes man," by flattering the other, or by attempting to emphasize additional values that he may be able to offer the other. He may resort to any one or all of these tactics. The research indicates that a low-status person will tend to resort to ingratiation to the extent that it is important that the person please the other and to the extent that he perceives that the other may be manipulated by such techniques. [See INTERACTION.]

We have confined this survey to a general overview and to a specific consideration of certain basic processes of perception and social influence. Extrapolation from these processes can be made to a number of problem areas that come within the domain of the social psychologist. The proliferation of the two-person situation generates social



structures in which each person's relationship to others in the structure is defined by behavioral expectations and rewards to be acquired by the person and by the rewards that he is expected to provide for others. He may enter into a relationship of mutuality with one or more other persons within the structure. In order for a structure to function effectively, certain norms of belief and conduct peculiar to the structure tend to develop. Situations arise in which it is necessary for one person to assume the responsibility for coordinating the behaviors and interactions of the others. These considerations have brought the study of leadership within the purview of the social psychologist. Social psychologists have also been involved in the study of large-scale organizations, in which subgroups are articulated into an over-all structure. [See ORGANIZATIONS.]

HAROLD B. GERARD

#### BIBLIOGRAPHY

- ALLPORT, FLOYD H. 1920 The Influence of the Group Upon Association and Thought. *Journal of Experimental Psychology* 3:159-182.
- ARONSON, ELLIOT; TURNER, JUDITH A.; and CARLSMITH, J. MERRILL 1963 Communicator Credibility and Communication Discrepancy as Determinants of Opinion Change. *Journal of Abnormal and Social Psychology* 67:31-36.
- ASCH, SOLOMON E. 1948 The Doctrine of Suggestion, Prestige and Imitation in Social Psychology. *Psychological Review* 55:250-276.
- ASCH, SOLOMON E. 1956 Studies of Independence and Conformity: I. A Minority of One Against a Unanimous Majority. *Psychological Monographs* 70, no. 9.
- BENNETT, EDITH B. 1955 Discussion, Decision, Commitment and Consensus in "Group Decision." *Human Relations* 8:251-273.
- BERKOWITZ, LEONARD 1957 Liking for the Group and the Perceived Merit of the Group's Behavior. *Journal of Abnormal and Social Psychology* 54:353-357.
- BREHM, JACK W.; and COHEN, ARTHUR R. 1962 *Explorations in Cognitive Dissonance*. New York: Wiley.
- BROWN, ROGER W. 1965 *Social Psychology*. New York: Free Press.
- BRUNER, JEROME S. 1957 Compatriot and Foreigner: A Study of Impression Formation in Three Countries. *Journal of Abnormal and Social Psychology* 55:253-260.
- CHAPMAN, DWIGHT W.; and VOLKMANN, JOHN A. 1939 A Social Determinant of the Level of Aspiration. *Journal of Abnormal and Social Psychology* 34:225-238.
- COHEN, ARTHUR R.; GREENBAUM, CHARLES W.; and MANS-SON, HELGE H. 1963 Commitment to Social Deprivation and Verbal Conditioning. *Journal of Abnormal and Social Psychology* 67:410-421.
- COOPER, EUNICE; and DINERMAN, HELEN 1951 Analysis of the Film, "Don't Be a Sucker": A Study in Communication. *Public Opinion Quarterly* 15:243-264.
- COOPER, EUNICE; and JAHODA, MARIE 1947 The Evasion of Propaganda: How Prejudiced People Respond to Anti-prejudice Propaganda. *Journal of Psychology* 23:15-25.
- DEUTSCH, MORTON 1958 Trust and Suspicion. *Journal of Conflict Resolution* 2:265-279.
- DEUTSCH, MORTON; and GERARD, HAROLD B. 1955 A Study of Normative and Informational Social Influences Upon Individual Judgment. *Journal of Abnormal and Social Psychology* 51:629-636.
- DI VESTA, FRANCIS J.; and MERWIN, JACK C. 1960 The Effects of Need-oriented Communications on Attitude Change. *Journal of Abnormal and Social Psychology* 60:80-85.
- FESTINGER, LEON 1954 A Theory of Social Comparison Processes. *Human Relations* 7:117-140.
- FESTINGER, LEON 1957 *A Theory of Cognitive Dissonance*. Evanston, Ill.: Row, Peterson.
- FESTINGER, LEON; and FREEDMAN, JONATHAN L. 1964 Dissonance Reduction and Moral Values. Pages 220-243 in Symposium on Personality Change, University of Texas, *Personality Change*. New York: Wiley.
- FESTINGER, LEON; and MACCOBY, NATHAN 1964 On Resistance to Persuasive Communications. *Journal of Abnormal and Social Psychology* 68:359-366.
- FESTINGER, LEON; SCHACHTER, STANLEY; and BACK, KURT (1950) 1963 *Social Pressures in Informal Groups: A Study of Human Factors in Housing*. Stanford Univ. Press.
- FESTINGER, LEON; and THIBAUT, JOHN 1951 Interpersonal Communication in Small Groups. *Journal of Abnormal and Social Psychology* 46:92-99.
- FESTINGER, LEON et al. 1952 The Influence Process in the Presence of Extreme Deviates. *Human Relations* 5:327-346.
- FREEDMAN, JONATHAN L.; and SEARS, DAVID O. 1965 Warning, Distraction and Resistance to Influence. *Journal of Personality and Social Psychology* 1:262-266.
- GERARD, HAROLD B. 1953 The Effect of Different Dimensions of Disagreement on the Communication Process in Small Groups. *Human Relations* 6:249-271.
- GERARD, HAROLD B. 1961 Disagreement With Others, Their Credibility and Experienced Stress. *Journal of Abnormal and Social Psychology* 62:559-564.
- GERARD, HAROLD B. 1965 Deviation, Conformity and Commitment. Pages 263-277 in Ivan D. Steiner and Martin Fishbein (editors), *Current Studies in Social Psychology*. New York: Holt.
- GEWIRTZ, JACOB L.; and BAER, DONALD M. 1958 The Effect of Brief Social Deprivation on Behaviors for a Social Reinforcer. *Journal of Abnormal and Social Psychology* 56:49-56.
- GREENSPOON, JOEL 1955 The Reinforcing Effect of Two Spoken Sounds on the Frequency of Two Responses. *American Journal of Psychology* 68:409-416.
- HARVEY, O. J.; and CONSALVI, CONRAD 1960 Status and Conformity to Pressures in Informal Groups. *Journal of Abnormal and Social Psychology* 60:182-187.
- HEIDER, FRITZ 1958 *The Psychology of Interpersonal Relations*. New York: Wiley.
- HOVLAND, CARL I. 1959 Reconciling Conflicting Results Derived From Experimental and Survey Studies of Attitude Change. *American Psychologist* 14:8-17.
- HOVLAND, CARL I.; CAMPBELL, ENID H.; and BROCK, TIMOTHY 1957 The Effects of "Commitment" on Opinion Change Following Communication. Pages 23-32 in Carl I. Hovland (editor), *The Order of Presentation in Persuasion*. New Haven: Yale Univ. Press.
- HOVLAND, CARL I.; HARVEY, O. J.; and SHERIF, MUZAFAER 1957 Assimilation and Contrast Effects in Reactions

- to Communication and Attitude Change. *Journal of Abnormal and Social Psychology* 55:244-252.
- HOVLAND, CARL I.; JANIS, IRVING L.; and KELLEY, HAROLD H. 1953 *Communication and Persuasion: Psychological Studies of Opinion Change*. New Haven: Yale Univ. Press.
- HOVLAND, CARL I.; and MANDELL, WALLACE 1952 An Experimental Comparison of Conclusion-drawing by the Communicator and by the Audience. *Journal of Abnormal and Social Psychology* 47:581-588.
- ITARD, JEAN M. G. (1801) 1932 *Wild Boy of Aveyron*. New York: Century. → First published as *De l'éducation d'un homme sauvage* . . .
- JANIS, IRVING L.; and KING, BERT T. 1954 The Influence of Role Playing on Opinion Change. *Journal of Abnormal and Social Psychology* 49:211-218.
- JONES, EDWARD E. 1964 *Ingratiation: A Social Psychological Analysis*. New York: Appleton.
- KELLEY, HAROLD H.; and SHAPIRO, MARTIN M. 1954 An Experiment on Conformity to Group Norms Where Conformity is Detrimental to Group Achievement. *American Sociological Review* 19:667-677.
- KELLEY, HAROLD H.; and WOODRUFF, CHRISTINE L. 1956 Members' Reaction to Apparent Group Approval of a Counternorm Communication. *Journal of Abnormal and Social Psychology* 52:67-74.
- KELLEY, HAROLD H. et al. 1962 The Development of Cooperation in the "Minimal Social Situation." *Psychological Monographs* 76, no. 19.
- KRECH, DAVID; CRUTCHFIELD, RICHARD S.; and BELLACHEY, EGERTON L. (1948) 1962 *Individual in Society: A Textbook of Social Psychology*. New York: McGraw-Hill. → First published as *Theory and Problems of Social Psychology*, by Krech and Crutchfield.
- LEVENTHAL, H.; and KAFES, P. N. 1963 The Effectiveness of Fear Arousing Movies in Motivating Preventive Health Measures. *New York State Journal of Medicine* 63:867-874.
- LEWIN, KURT (1947) 1958 Group Decision and Social Change. Pages 197-211 in *Society for the Psychological Study of Social Issues, Readings in Social Psychology*. 3d ed. New York: Holt.
- LINDZEY, GARDNER (editor) (1954) 1959 *Handbook of Social Psychology*. 2 vols. Cambridge, Mass.: Addison-Wesley. → Volume 1: *Theory and Method*. Volume 2: *Special Fields and Applications*.
- MCGUIRE, WILLIAM J. 1964 Inducing Resistance to Persuasion: Some Contemporary Approaches. Volume 1, pages 191-229 in *Advances in Experimental Social Psychology*. Edited by Leonard Berkowitz. New York: Academic Press.
- MERTON, ROBERT K.; and KITT, ALICE S. 1950 Contributions to the Theory of Reference Group Behavior. Pages 40-105 in Robert K. Merton and Paul F. Lazarsfeld (editors), *Continuities in Social Research: Studies in the Scope and Method of The American Soldier*. Glencoe, Ill.: Free Press.
- MILLER, NORMAN; and CAMPBELL, DONALD T. 1959 Recency and Primacy in Persuasion as a Function of the Timing of Speeches and Measurements. *Journal of Abnormal and Social Psychology* 59:1-9.
- MILLS, JUDSON 1958 Changes in Moral Attitudes Following Temptation. *Journal of Personality* 26:517-531.
- NEWCOMB, THEODORE M. (1943) 1957 *Personality and Social Change: Attitude Formation in a Student Community*. New York: Dryden.
- NEWCOMB, THEODORE M. 1963 Persistence and Regression of Changed Attitudes: Long-range Studies. *Journal of Social Issues* 19, no. 4:3-14.
- NEWCOMB, THEODORE M.; TURNER, RALPH H.; and CONVERSE, PHILIP E. 1965 *Social Psychology: The Study of Human Interaction*. New York: Holt.
- PROSHANSKY, HAROLD; and SEIDENBERG, BERNARD (editors) 1965 *Basic Studies in Social Psychology*. New York: Holt.
- RAZBAN, GREGORY 1940 Conditioned Response Changes in Rating and Appraising Sociopolitical Slogans. *Psychological Bulletin* 37:481 only.
- SAMELSON, FRANZ 1957 Conforming Behavior Under Two Conditions of Conflict in the Cognitive Field. *Journal of Abnormal and Social Psychology* 55:181-187.
- SCHACHTER, STANLEY 1951 Deviation, Rejection, and Communication. *Journal of Abnormal and Social Psychology* 46:190-207.
- SCHACHTER, STANLEY 1959 *The Psychology of Affiliation: Experimental Studies of the Sources of Gregariousness*. Stanford Studies in Psychology, No. 1. Stanford Univ. Press.
- SECORD, PAUL F.; and BACKMAN, CARL W. 1964 *Social Psychology*. New York: McGraw-Hill.
- SHERIF, MUZAFAER 1935 A Study of Some Social Factors in Perception. *Archives of Psychology* 27, no. 187.
- SIDOWSKI, JOSEPH B.; WYCKOFF, L. BENJAMIN; and TABORY, LEON 1956 The Influence of Reinforcement and Punishment in a Minimal Social Situation. *Journal of Abnormal and Social Psychology* 52:115-119.
- SMITH, M. BREWSTER; BRUNER, JEROME S.; and WHITE, R. W. 1956 *Opinions and Personality*. New York: Wiley.
- STAATS, ARTHUR W.; and STAATS, CAROLYN K. 1958 Attitudes Established by Classical Conditioning. *Journal of Abnormal and Social Psychology* 57:37-40.
- STEINER, IVAN D.; and FISHBEIN, MARTIN (editors) 1965 *Current Studies in Social Psychology*. New York: Holt.
- WALSTER, ELAINE; and FESTINGER, LEON 1962 The Effectiveness of "Overheard" Persuasive Communications. *Journal of Abnormal and Social Psychology* 65:395-402.
- WEISS, WALTER 1957 Opinion Congruence With a Negative Source on One Issue As a Factor Influencing Agreement on Another Issue. *Journal of Abnormal and Social Psychology* 54:180-166.

## SOCIAL ROLE

See under ROLE.

## SOCIAL SCIENCE FICTION

The creation of imaginary worlds for the purpose of exploring the potential development of society represents one of the oldest traditions in literature. The classic list begins with Plato's *Republic* and includes Bacon's *New Atlantis*, More's *Utopia*, Campanella's *City of the Sun*, Swift's *Gulliver's Travels*, and Butler's *Erewhon*. Bellamy's *Looking Backward* (1888) ushers in the modern era. Subsequent examples are Zamiatin's *We* (1925),



Huxley's *Brave New World* (1932), and Orwell's 1984 (1949), which have been described as "black utopias," or "dystopias."

A common characteristic of this literature is an emphasis on the social landscape rather than the individual protagonist, on the articulation of artificial models of society rather than interaction between fictional counterparts of people in real life, on projections from reality rather than variations of actual social situations. Articles elsewhere in this encyclopedia deal with other aspects of this genre. [See LITERATURE, article on POLITICAL FICTION; UTOPIANISM.] This article focuses on a more recent variant of the genre: the novels, novelettes, and short stories that have proliferated since World War II and which can be grouped under the rubric "social science fiction."

### Scope

The term "social science fiction" can usefully be employed to identify narratives that extrapolate from current social science concepts in order to predict or speculate about the future shape of society. This new genre of fiction stems from two literary traditions: the modern utopias and dystopias mentioned above and classic science fiction, whose early practitioners were Jules Verne and H. G. Wells and which continues to proliferate. In fact, social science fiction is a contemporary development of classic science fiction: it involves the same imaginative leaps into the future, it uses some of the same stylized conventions (time travel, interplanetary explorations), props (spaceships, robots), and characters (aliens, androids), but only as incidental backdrops to a new category of concerns. It differs from classic science fiction in two important respects. First, whereas classic science fiction is concerned with predicting the shape of the physical world through imagining the favorable and adverse potentialities of the physical sciences, the organizing concept of social science fiction is speculation about the future of society through projecting potential innovations in the knowledge and techniques of the social sciences. Second, social science fiction takes the technological innovations of the future more or less for granted and focuses on their social consequences.

Social science fiction is occasionally produced by known writers of more orthodox literature; the bulk of it, however, comes from the pens of immensely prolific writers who specialize in science fiction. The knowledge of the social sciences displayed by most of these writers is understandably naive and superficial since it is obviously gleaned secondhand from popularizations of the various disciplines.

Much of social science fiction is so inept in style and plot that it hardly merits consideration as literature, even according to the most tolerant critical standards. Much of it is published between the same lurid magazine and paperback covers as is classic science fiction and is announced in the same penny-dreadful blurbs. As a result, it has shared the same fate of being dismissed by literary critics as a minor, cultist offshoot of traditional literature, of interest only to an intellectually limited circle of *aficionados*. To the social scientist, however, the emergence of a vast body of literature on the periphery of his academic vision, which incorporates—however inaccurately—his specialized perspectives, is of more than casual interest.

### Significance

Social science fiction is noteworthy on several levels: first, as a cultural phenomenon; second, as a medium of social commentary and criticism; and third (perhaps parenthetically) as a pleasant counterpoint to some of the more pedantic complacencies of the social scientists.

As a cultural phenomenon, social science fiction provides evidence of the growing impact of the social sciences on popular culture, just as the older genre of classic science fiction documented popular interest in the successive stages of the technological revolution of the nineteenth and twentieth centuries. Between Jules Verne's 1870 submarine and the predictions of nuclear fission and space travel of the 1930s and 1940s, radical changes took place in the perspectives and expectations of the average man—changes that were in part wrought by innovations in the physical sciences. With the closing of the gap between imaginative pseudo science and almost prosaic reality, with the advent of atomic power plants, satellite traffic jams, and space walks, popular interest in the fictionalized potential of the physical sciences has become jaded. There is mounting evidence that it is being superseded by a fascination with the potential applications of the social sciences. The form this interest takes and the purposes to which writers of social science fiction bend—or distend—current knowledge in these fields constitute the subject matter of this article.

That social science fiction is an instrument for social commentary and criticism, ranging in tone from earnest concern to outrageous parody, from macabre warnings to comic alarums, is inherent in its subject matter. Three major themes are discernible both in the plethora of wildly imaginative worlds of science fiction writers and in the works of more sober and orthodox writers. The first of these centers on the social aftermath of some

cosmic disaster. These "post-doomsday" stories begin with the collapse of civilization, the agent of destruction being a nuclear war or a natural catastrophe, such as a drastic change in the weather, an epidemic, or a massive ecological imbalance. The events of the cataclysm provide dramatic interest, but they are incidental to the main theme of the story, which deals with the disintegration of social institutions, the breakdown of mechanisms for social control, and the emergence of new leaders, a new social system, and a new culture.

The second major theme consists of serious extrapolations from (or gleeful projections *ad absurdum* of) trends in contemporary society. These range from uncontrolled population growth or total depletion of crucial natural resources to such cultural aberrations as mass drug addiction or psychosis, the rampant development of fads, ultra-conformism, or the total commercialization of society.

The third major theme is the exploration of the effects upon individuals of life in unbalanced or truncated social systems. Novels in this category are concerned with what *might* be if one element in society (such as teen-agers) should seize control, or one particular technique of planning be allowed to dominate (as in the total planning for consumption), or one particular set of values (such as hedonism) becomes an enforced mode of behavior. The settings are skewed models of society built by writers with an experimental bent whose point of departure is the question "What if . . .?" and whose concern is with the processes and techniques of social manipulation.

Social science fiction is essentially moralistic. The writers of post-doomsday stories ask what happens to ethical and cultural values in the midst of social disorganization and chaos. Writers with a satirical purpose fabricate unidimensional societies top-heavy with absurd exaggerations of today's mores or aspirations for the express purpose of watching them topple. The creators of Orwellian nightmares exalt the struggle for individual freedom against the enslavement of will and intellect and sound warnings of the stultifying effects of rigid one-class or multiclass stratification systems. Like the insistent hum of the diabolical machines of classic science fiction, there is in social science fiction an undercurrent of distrust of the potentialities of unbridled social scientism—of the potential menace of techniques of indoctrination, planned acculturation, population control, and the use of statistical averages as required behavioral norms. The "mad scientist" has been replaced by a species of more plausible villain: the smooth organization

man, the amoral advertising tycoon, the cynical censor, the bland but dictatorial bureaucrat, the voracious practitioner of motivation research, the psychiatrist-in-power and his squad of ubiquitous "medicops," the amiable psychologist in charge of "preconditioning." The death ray has largely been superseded by the giant computer, and even the spying telescreen used by Big Brother in 1984 has been supplanted by the television wall, the controlled alternative to "real" social participation—"the family" as it is called in Ray Bradbury's *Fahrenheit 451* (1953). Much of this distrust of the role and the technology of the social sciences, it is well to remember, is simply the result of inadequate understanding, as well as a reflection of a fundamentally populist bias against all science.

### Types of social science fiction

With this outline of the scope and significance of social science fiction in mind it is desirable to turn to some specific examples. Some represent "purer" specimens than others in the sense that they draw upon one discipline, or one topic in a discipline (such as learning theory), for the setting and action. Others, particularly the panoramic novels, chronicle the death and rebirth of civilization. The pseudo histories of the future—such as Walter M. Miller's *A Canticle for Leibowitz* (1959), which describes a feudal theocratic society that is the preserver of civilization between atomic holocausts—borrow concepts from all the social sciences. Nevertheless, for present purposes it is possible to group the examples selected according to one or another disciplinary label. Since many of the works of social science fiction described are concerned with the subject matter of political science, that is, with the nature and locus of power and authority in society, and since several examples of political science fiction are described elsewhere, political science is not included as a separate category [see LITERATURE, article on POLITICAL FICTION].

**Anthropology.** The popularization of anthropology and its subdisciplines has provided new insights and perspectives to an entire generation of writers of fiction. In particular, the concept of culture as a holistic approach to the varieties of human life has had an impact perhaps comparable to that of psychoanalysis.

A reflection of this impact is found in those novels and short stories whose main preoccupation is with the components of culture; with the dynamics of culture change under stress; with the rise and development of new cultures on earth under primitive conditions following some trau-



matic event which destroys contemporary civilization; or with the description of different cultures in other worlds in some distant future and the attendant problems of communication and contact between cultures.

*Earth Abides* by George R. Stewart (1949) shows evidence of having been inspired by the concepts of cultural anthropology. Two themes permeate this story, which concerns a handful of survivors in a world that has been nearly depopulated by a mutated virus. First, the founder and leader of "The Tribe"—as the group is called—fails in his attempt to teach the children who are born into the group the values and customs of the old civilization. Each society must develop its own cultural symbols, Stewart seems to say, if it is to survive and grow. A miner's hammer owned by the leader, for example, becomes the focal symbol of his authority. Second, the emergence of the new culture illustrates the conflict between the individual and society. When a straggler appears on the scene and it is discovered that he has a venereal disease, the elders vote to hang him in order that The Tribe may survive. After the execution, the leader reflects on the meaning of what has been done:

This was an end, and this was also a beginning. It was the end of those twenty-one years when they had lived, now he thought, . . . as it might have been in some old Garden of Eden. . . . In the past, there had been only a little group of people, scarcely more than an overgrown family. In the future, there would be the State. . . . And now the first act of the State, its originating function, had been to bring death. (1949, p. 247 in the Ace edition)

Chad Oliver, an anthropologist, understandably provides the most knowledgeable and integral examples of fiction which incorporates and elaborates upon culture concepts. In the novel *Shadows in the Sun* (1954a), Oliver extrapolates both from the methodology of cultural anthropology and the concept of culture patterns. The narrative concerns the experiences of the hero (a cultural anthropologist) as he analyzes the data obtained from a community study of an American small town. The data do not produce the expected correlations; there seems to be something strange about the culture which was not detectable during the field work. Gradually the explanation emerges: the town is populated solely by aliens from another planet who have slowly bought up the property and settled there in order to have a base camp on earth. They have successfully posed as earthlings until culture-pattern analysis unmask them.

Oliver's short story "Rite of Passage" (1954b)

also focuses on the concept of cultural configuration. Three survivors of the crew of an interplanetary exploration expedition, one of them the official anthropologist, land on a planet of another solar system. They are welcomed by the inhabitants, whose culture is apparently primitive and pastoral. After a brief sojourn among the Nern, during which the anthropologist is taught the language by the village shaman, the other crew members are eager to introduce the basic elements of technology. The anthropologist, however, deduces correctly that the Nern have eschewed technology by deliberate option, by "editing" their culture "down to essentials," i.e., the goal of "survival with maximum integration, cohesiveness of function, individual fulfillment, constant challenge, and peace" ([1954b] 1966, p. 282). They are "masters of the culture process," experts at introducing their own ideas in other cultures at a distance through psychology, hypnosis, and adroit appeals, at turning enemies into allies, and destroying them.

Linguistics is also a favored theme in anthropological fiction, particularly those aspects that deal with the relationship of language to culture. What fascinates writers of social science fiction is the thesis that language can control behavior. Perhaps the best example is in 1984, a novel which is primarily political science fiction. *Newspeak*, the language devised by the totalitarian regime described by Orwell, is one of the state's major instruments of thought control. Its purpose, as Orwell describes it in the appendix to the novel, is "not only to provide a medium of expression for the world-view and mental habits proper to the devotees of Ingsoc [English socialism], but to make all other modes of thought impossible" (1949, p. 246 in the 1961 edition).

Jack Vance's *The Languages of Pao* (1957) is another example of linguistic fiction. The story, set far in the future, centers on the thesis that language creates cultural values and thus can be deliberately used to create new social systems. The ruler of an "acquiescent" society on the planet Pao obtains technical-assistance experts from another planet to help him defend Pao against its enemies. The visiting superscientists are anthropologists and linguists, and they transform Pao into an aggressive caste society by creating and inculcating three separate languages: one for the military, with a syllabary "rich in effort-producing gutturals and hard vowels"; one for the industrialists and technicians, with an extravagantly complicated but logical grammar and "elaborate rules of accordance"; and one for the traders, with "emphatic

number-parsing, elaborate honorifics . . . to facilitate ambiguity, a syntax of reflection, reinforcement and alternation to emphasize the analogous interchange of human affairs" ([1957] 1958, p. 57).

Another subdiscipline of anthropology—archeology—has provided inspiration for writers of the science fiction school. The reconstruction through archeological techniques of vanished civilizations on distant planets has introduced a new twist to the so-called space operas of an earlier period. The examples are numerous and endlessly fanciful. In fact, the physicist-biochemist Leo Szilard wrote a parody that seems aimed both at archeologists and archeological fiction. Entitled "Grand Central Terminal" (1952), it solemnly describes the efforts of "alien" archeologists from space to reconstruct the image of earth culture from the ruins of the station and its public facilities.

**Sociology.** The pervasive influence of the concepts and terminology of sociology is reflected in works of social science fiction that explore new forms of social stratification or new mechanisms of social control or which are imaginative projections of such present-day social pathologies as overpopulation, drug addiction, organized crime, juvenile delinquency, and the manipulation of behavior through the mass media.

Although a popular form of sociological science fiction is the dystopia, in the tradition of 1984, both the character of the menace and the objectives of the protest are different from those described by Orwell. The contemporary sociological version of the dystopia is concerned not with the problem of oligarchy but with the ultimate triumph of a mindless mass mentality. Two novels, *The Thirty-first Floor* by the Swedish writer Peter Wahlöö (1964) and *Fahrenheit 451* by the American Ray Bradbury, are cases in point. Both novels describe static, egalitarian societies of the future that are dominated by a mass culture, in which independent thought is stifled and conformity to the bland average in behavior is enforced. The values of these societies are expressed and promulgated by the homogenized mass media. Reduced to the lowest common denominator of emotional and intellectual appeal, the press and television eschew all controversy, offend no one, and level all aspirations to a moronic hedonism. The institutional means of control is censorship. In Wahlöö's world the censorship is unofficial, unobtrusive, but nonetheless pervasive. Private enterprise and government are intertwined and opinion is controlled by a democratic combine of the press, the trade unions, and the political parties. A vast publishing

concern shields the individual's peace of mind and security ([1964] 1967, p. 37) from all information about discontent (suicide, deviance, drunkenness) or negative frightening realities (crime, pollution, overcrowding). It is "the majority of people . . . [who] act as censors" (*ibid.*, p. 175); society protects itself. In Bradbury's dystopia, censorship is overt: the agents of control are the firemen, whose assigned duty is not to put out fires but to destroy books, which are outlawed as incendiary threats to the mass mentality (451° F. is the temperature at which book paper is said to burst into flame). As the Firechief explains to the fireman hero Montag, who has begun to read clandestinely the books he is supposed to burn: "It didn't come from the Government down. There was no dictum, no declaration, no censorship to start with, no! Technology, mass exploitation, and minority pressure carried the trick" (1953, p. 53). But the social process became an ideology:

We must all be alike. Not everyone born free and equal, as the Constitution says, but everyone *made* equal. . . . then all are happy. . . . Don't give them any slippery stuff like philosophy or sociology to tie things up with. That way lies melancholy. Any man who can take a TV wall apart and put it back together again . . . is happier than any man who tries to slide-rule, measure, and equate the universe. . . . (*ibid.*, pp. 53, 56)

True to form, both Wahlöö and Bradbury exalt the rebel against the rationalized society. In *The Thirty-first Floor* a dissident editor, tricked with other intellectuals into collaboration with the "concern," threatens to blow up its headquarters. In *Fahrenheit 451* Montag joins the underground—people living in the hills and backwaters who also burn books but memorize them first. Each person in fact is a book ("I am Plato's *Republic*" is an introduction). In this way the oral traditions of man's preliterate past enable civilization to survive until a nuclear holocaust, beginning as the book ends, destroys the centers of anti-intellectualism.

In *The 10th Victim*, a humorous novel about the future by Robert Sheckley (1965), another social process is described. In the twenty-first century, a society sick of war has finally evolved a functional substitute for aggression: the legalization of murder through the institution of the Hunt. The Hunt is described as an international contest, in which participation is voluntary. In all the capitals of the civilized world, giant computers randomly select pairs of antagonists who are licensed as "hunter" and "victim." Prize money is awarded to the victors; the hunters and victims alternate roles until one of them reaches the limit of ten kills. Intended for



As a result of aggression the threat to other neighborhood  
is unknown whereas other neighborhood is not  
aggression, p. 20

Several developmental milestones in the child language domain indicate that the child is ready to learn a second language. These include the child's ability to understand a wide range of language input. Most infants learn to understand and discriminate between all the linguistic forms that are presented to them in the first 12 months of life. They begin to understand the meaning of words and to use language to communicate with others. They also begin to understand the meaning of words and to use language to communicate with others. They also begin to understand the meaning of words and to use language to communicate with others.

[illegible]

Two kinds of very important information may be obtained from the analysis of the following statements. When the number of all items is known, it is possible to determine the value of the items. The value of the items is the sum of the items. The value of the items is the sum of the items.

[illegible]

A third source of the gap is differences in the way that individuals in the two countries interpret the question. In the United States, the question is interpreted as asking for the number of people who are currently in the military. In the United Kingdom, the question is interpreted as asking for the number of people who have ever served in the military.

On the other hand, it is also true that the government has been successful in its efforts to control inflation and stabilize the economy. The government's policies have helped to bring about a period of relative stability and growth.

[illegible]

**Comments:** This book is an amazing glimpse into the world of a young woman, performed in a clear, accessible manner. It is the only volume of the series that speaks of her childhood, as *The Glass Heart* is devoted to her years as a sheltered, pampered princess, and *An Extraordinary Girl* is devoted to the turbulent adult years, and the journey toward peace, wisdom, and forgiveness. *Unsettled* is the only book in the series that speaks of her years as a sheltered, pampered princess, and *An Extraordinary Girl* is devoted to the turbulent adult years, and the journey toward peace, wisdom, and forgiveness. *Unsettled* is the only book in the series that speaks of her years as a sheltered, pampered princess, and *An Extraordinary Girl* is devoted to the turbulent adult years, and the journey toward peace, wisdom, and forgiveness.

THE UNIVERSITY OF CHICAGO PRESS





...the ... of ...

...the ... of ...

...the ... of ...

...the ... of ...

...the ... of ...

...the ... of ...

...the ... of ...

...the ... of ...

...the ... of ...

...the ... of ...

...the ... of ...

...the ... of ...

...the ... of ...

...the ... of ...

...the ... of ...

...the ... of ...

...the ... of ...

...the ... of ...

...the ... of ...

...the ... of ...

...the ... of ...

...the ... of ...

...the ... of ...

...the ... of ...

...the ... of ...

...the ... of ...

...the ... of ...

...the ... of ...

[illegible][illegible]



will be found in AREA STUDIES, BEHAVIORAL SCIENCES, CONFERENCES, ETHICS, article on ETHICAL ISSUES IN THE SOCIAL SCIENCES, INFORMATION STORAGE AND RETRIEVAL, particularly the articles ON REFERENCE MATERIALS AND BOOKS and BIBLIOGRAPHIC ISSUES IN THE BEHAVIORAL SCIENCES, SCIENCE. Problems of financial support are reviewed in FOUNDATIONS, RESEARCH AND DEVELOPMENT, article on FINANCING SOCIAL RESEARCH. Material on the utilization of the social sciences and of social scientists will be found in POLICY SCIENCES, PROFESSIONS, SOCIAL PROBLEMS, TEACHING, UNIVERSITIES. For an overview of the coverage in this encyclopedia of specific fields of study, see ANTHROPOLOGY, ECONOMICS, GEOGRAPHY, HISTORY, LAW, POLITICAL SCIENCE, PSYCHIATRY, PSYCHOLOGY, SOCIOLOGY, STATISTICS

### SOCIAL SECURITY

See AGING, article on ECONOMIC ASPECTS, MEDICAL CARE, article on ECONOMIC ASPECTS, UNEMPLOYMENT INSURANCE. For coverage of other governmental and private assistance, see the articles listed under SOCIAL WELFARE

### SOCIAL STATISTICS

See CENSUS, GOVERNMENT STATISTICS, SAMPLE SURVEYS, STATISTICS, VITAL STATISTICS

### SOCIAL STATUS

See STATUS, SOCIAL.

### SOCIAL STRATIFICATION

See STRATIFICATION, SOCIAL.

## SOCIAL STRUCTURE

### 1 THE HISTORY OF THE CONCEPT OF SOCIAL STRUCTURAL ANALYSIS

Edmund R. Leach  
Stanley H. Udy, Jr.

#### THE HISTORY OF THE CONCEPT

In academic work the evaluation of concepts depends upon their use by specialists. In the decade following World War II the concept of "social structure" became extremely fashionable among social anthropologists and at times attained even extreme prominence. It should be apparent to anyone who has followed developments of recent years that the concept has not been served by pursuing all these ramifications of usage. Instead, this article evaluates the meanings of the concept adopted by George Murdock in America, A. R. Radcliffe-Brown and various of his followers in Britain, and

Claude Lévi-Strauss in France. The usages of other writers may be considered as modifications of one or other of these relatively extreme positions. Although nationality is relevant here, it is not decisive; there are British anthropologists whose position is close to that of Lévi-Strauss just as there are Americans whose postulates are very similar to those of Radcliffe-Brown, but "structuralist anthropology" has a different connotation in France and in Britain, and the two meanings should not be confused.

Early uses of the concept. Although the fashionable use of the term is a post-1945 phenomenon, the concept of "social structure" has a long history, which may be briefly reviewed. In its original English meaning the word "structure" referred to building construction, but by the sixteenth century it was also being used to refer to the interrelations between the component parts of any whole. It was in fact widely used in this sense in anatomical studies, which were in active development at that period. Its further extension from anatomy to sociology, though delayed for several centuries, was a logical corollary of the very general use of organic analogies by political philosophers. Hobbes does not actually use the expression "social structure" in his *Leviathan*, but his conception of the state as an artificial organism, in which the function of each component institution and office holder is carefully distinguished, is one to which the term might very well have been applied even in the seventeenth century. The explicit idea that the study of social structure should be an objective for sociological enquiry seems to be due to Spencer (1858), who refers to "the Inductions of Sociology—general facts, structural and functional as gathered from a survey of societies and their changes."

The fact that Spencer brought the terms "structure" and "function" into direct association shows that he had an anatomical image in mind. This image appears again in the work of Durkheim, who was the source of Radcliffe-Brown's ideas. As early as 1910 the latter gave lectures on Durkheimian sociology under the title "Social Structure." Provided that we accept the organic analogy, there is no difficulty about understanding what these authors mean. Society is treated as a kind of living creature, the parts of which can be dissected and studied. The "social structure," then, is the mesh of mutual positions and interrelations in terms of which the interdependence of the component parts may be described, the "function" of any part is the way it operates so as to maintain the total system in good health. But as soon as we ask such questions as Is society really an object

ism? Where do we discern the boundary between one society and the next? How do we distinguish between a society that is in good health and one that is in a pathological condition?, the extreme artificiality of the model becomes apparent.

A less clearly defined strand of thought, which is free of organic overtones, is to be discerned in early Marxist literature. This is relevant for anthropology because of the close dependence of Engels (1884) upon Morgan (1877). Marx (1859) had written of the relations of production as constituting "the economic structure (*Struktur*), the real basis (*Basis*) on which is erected a juridical and political superstructure (*Oberbau*) and to which correspond the forms (*Formen*) of the determined social conscience." Marx's metaphor here is plainly that of a building, not of an organism, but the notion of structure is not sharply distinguished from a variety of other concepts. Elsewhere there are references to political, juridical, religious, and philosophical systems—in contexts where the word "system" is indistinguishable from the above uses of "structure," "superstructure," and "form" (Lefebvre 1962). More recently, sociologists have added such variants as "infrastructure," "macrostructure," "microstructure" on an apparently arbitrary basis. Engels (1884) specifically refers to Morgan's "Stage of Savagery" as one in which "the social structure is based on ties of consanguinity," but we need not suppose that he had any very clearly defined concept in mind, such expressions as "social order," "social system," or "social form" would have served just as well. On the other hand, if one were to seek a candidate for the first anthropological work which may be considered a "study of social structure" (in any of the three senses under discussion), the most obvious candidate is certainly Morgan's work (1871), where the title word is *Systems*. The terms "system" and "structure" are not always synonymous in anthropological writing, but they are often indistinguishable.

Murdock. The meaning which Murdock attaches to the expression "social structure" may be gathered from the items listed below. His principal concerns are the ethnographic facts and the taxonomic classification of societies on the basis of manifest, readily discernible characteristics. The Human Relations Area Files, located at Yale, which were originated by Murdock and which have remained his principal research tool, are an attempt to compare societies systematically by analyzing written source materials (of very varied quality and origin) according to a strictly uniform procedure. It is presupposed that the tribal names which appear in anthropological monographs correspond

to fixed, readily distinguishable social entities (societies, tribes, cultures), which are fully described by listing the sum of their attributes. Murdock has drawn up a list of such possible attributes (Yale University 1938), and the immediate purpose of the Area Files (HRAF) analysis is to determine whether each listed attribute does or does not occur in the particular "society" under investigation. The taxonomy established by Murdock (1949) depends primarily on varieties of kinship organization. Insofar as Murdock uses an organic analogy, it is that of a dissected model laid out in a showcase, there is no attempt to envisage society as a living entity capable of growth. The taxonomy is based upon statistical correlation rather than functional analysis. The similarity of society A to society B is a factor measured by the number of instances in which the two societies share the same attributes as measured by the HRAF scheme of analysis. The ultimate objective is historical. In botany, Linnaeus' taxonomy, based on classification by attributes, was later adapted into a scheme for tracing lines of evolution on a phylogenetic basis. Murdock claimed that his social taxonomy could likewise lead to evolutionary inferences (1949, chapter 8 and appendix). Critics reject the view that social aggregates are closely comparable to botanical species and argue that Murdock's view of cultural attributes is misleading. Concepts such as "patrilineal descent," "matrilocal residence," and so forth do not refer to characteristics which are present or absent in any clear-cut way in any particular case. Murdock's use of statistics shows that he considers each of his listed cultural attributes intrinsically independent of any other. This is completely at variance with the assumptions of the functionalist anthropologists, who presuppose an intrinsic interdependence between the parts of the whole.

Despite his use of the term "social structure" in book titles, Murdock's attitude has been ambivalent. He denigrates the concept as "static" and "sterile" (1955). He argues that emphasis on "structure" should be replaced by an emphasis on "process"; "The Linnaean system," he remarks, "only became something living after Darwin had discovered the process of variation and natural selection." Vogt (1960) makes a similar distinction, contrasting the "static" notion of structure (which he sees as a characteristic preoccupation of British anthropologists) with the "dynamic" notion of process favored by American scholars. It is not clear to the present writer how "social process" in Vogt's sense can differ from "history," and it appears that all those who write in this manner are in some meas-



ure historical determinists who believe that "laws" of social change can be discovered if we only look hard enough. The assumption that synchronic studies are necessarily "static" while diachronic studies are "dynamic" is not borne out by the facts, and the identification of this antithesis with a contrast between structure and process is meaningful only if structure is restricted to the formal and static taxonomic sense adopted by Murdock. For example, Fortes' "Time and Social Structure: An Ashanti Case Study" (1949) and Goody's collection of essays (1958), which derive from it, are all based on synchronic studies; they are all studies of social structure (in the British but not in Murdock's sense), but none of them is concerned with static phenomena.

**Radcliffe-Brown and the British school.** Murdock's use of the word "structure" implies either a building analogy ("society is an assemblage of independent bricks") or a dead organic model dissected for demonstration. In contrast, Radcliffe-Brown presumes that society may be compared to a living organism or a working mechanism. For Radcliffe-Brown, a society has a life of its own: it is not an object so much as a creature, so that the study of structure—that is, the interdependence of the component parts of the system—is indissolubly linked with the study of function, or how the component parts of the system "work" in relation to each other and to the whole.

Radcliffe-Brown's explicit formulation of this view and its application to anthropological studies are post-1931 developments. Radcliffe-Brown's *A Natural Science of Society* (1957) shows how he was thinking about these matters in 1937, when it was written. Bateson (1936) illuminates directly the parting of the ways between the cultural analyses of Malinowski and Benedict, on the one hand, and the structural view favored by Radcliffe-Brown, on the other. The structuralism of Radcliffe-Brown was developed as a reaction to the lack of precision shown by Malinowski in his handling of the concept of culture. Warner's *A Black Civilization* (1937), written under Radcliffe-Brown's tutelage, is the earliest monograph which is unambiguously structuralist in style in this sense.

Although Radcliffe-Brown's views of social function depend upon a holistic view of social structure, his British followers have not, in general, concerned themselves with the analysis of total working systems. Radcliffe-Brown himself was engaged in comparative work and theoretical generalization, and he encouraged his followers to concentrate on those aspects of a social system which display clear-cut formal characteristics and which

reflect to a minimal degree the interplay of personal rivalries. A characteristic sample of such studies may be found in Radcliffe-Brown (1924–1949), Fortes and Evans-Pritchard (1940), Radcliffe-Brown and Forde (1950), Evans-Pritchard (1940), and Fortes (1945). The authors concern themselves with only one formal aspect of society at a time—political structure, kinship structure, ritual organization, as the case may be. They display this system as a set of rules (jural obligations); they discuss the mutual interdependence of the rules and the fit between the society and its environment (ecology). But there is a tendency to play down any discrepancy between rule and practice. As a consequence of this, these studies show a marked neglect of economic organization.

**Firth and other British structuralists.** In contrast, other British writers such as Raymond Firth and A. I. Richards took over functionalist ideas (in a rather different form) from Malinowski but rejected Radcliffe-Brown's thesis that a system of jural regularities—a formal "structure"—may always be discerned behind the day-to-day operations of social life that an anthropologist observes in the field. It is notable that Firth and Richards pay much greater attention to economics than do the British structuralists and have consistently stressed the importance of the individual acting in his own interest as against the importance of Radcliffe-Brown's "social person," whose actions are fully defined by the rules which pertain to his social situation. A terminological detail calls for attention here. For Radcliffe-Brown and his followers the expression "social organization" is, in most cases, a synonym for "social structure," but Firth has for many years used "social organization" as a polar concept by which he seeks to emphasize the discrepancy between what actually happens and what a formal study of the rules might lead us to expect (see Firth 1964). The two viewpoints are complementary, and the tendency during the last few years has been for British writers to move closer to Firth's position (e.g., compare Forde 1950 with Forde 1963; both papers discuss the same aspects of Yakö social structure, but the former, written directly under Radcliffe-Brown's influence, lays far more stress on the formal aspects of the rules as distinct from their practical application).

**Weberian theory.** Over the years, British structuralism has been considerably modified by the assimilation of ideas derived from the sociology of Max Weber into a theoretical framework which was originally derived from Durkheim. Linton (1936) seems to have been the first anthropological author to make explicit use of the Weberian con-

cepts of status and role, although it seems possible that Linton himself was influenced by Radcliffe-Brown (then teaching in Chicago). A full-scale development of Weberian role theory as a tool of anthropological analysis is to be found in Nadel's *The Theory of Social Structure* (1957). Other Weberian concepts which have been extensively utilized by British structuralists are given below.

(1) There is the notion of "ideal types," which has sometimes been assumed to depend upon a discrimination similar to that by which Firth distinguished social structure from social organization. (2) There is the threefold typology of "traditional," "bureaucratic," and "charismatic" leadership. Weber's notion of the charismatic leader has proved particularly rewarding in the analysis of such phenomena as Melanesian cargo-cults, and an extensive literature discusses how far the structure of government in African states of the "traditional" type may be compared with or distinguished from "rational bureaucracy" of the western European type as understood by Weber (e.g., Fallers 1956). (3) Weber's concept of *Verband* was translated into English as "corporate group," thus suggesting an exact equivalence with the concept of "corporation," which has a sophisticated technical significance in English jurisprudence and which became assimilated into the language of British anthropology through its treatment by Maine (1861). In a penetrating summary of a large body of writing, Fortes (1953) pointed out that a crucial characteristic of the British structuralist studies of the interplay of political organization and kinship was that they treated unilineal descent groups (clans, lineages) as corporate groups in Weber's sense and that on this account a substantial body of Weberian theory became directly applicable to anthropological analysis.

It is notable that anthropological structuralism in the British sense has achieved its most convincing results when applied to social systems in which a segmentary structure of unilineal descent groups provides the over-all framework within which political activity takes place. In societies which lack unilineal descent groups and where, consequently, the Weberian categories are harder to apply, structuralist analysis has proved markedly less successful. As a by-product of this failure of theory, some writers have tried to prove that societies which lack unilineal descent groups are really just the same as those which possess them (Davenport 1959), the apparent object of this exercise being to show that Weberian categories can be applied with success even to "nonunilineal" descent groups. In the present writer's opinion this is a fallacy.

The ensuing debate has led to a clearer definition of various standpoints. The following summarizes what seems to be the present position of Fortes: Society is to be viewed as a system of corporations in Maine's sense. The corporations may be perpetual groups (corporations aggregate) or perpetual offices (corporations sole). Each such corporation has associated with it a set of jural rules which specify *inter alia* what the corporation possesses (its "estate"), who the members are ("the principle of recruitment"), how leadership within the corporation is exercised, what the formal rights and obligations of the corporation as a whole are vis-à-vis other like corporations, what the formal rights and obligations of the subunits of a corporation are vis-à-vis one another. The ultimate units of any such segmentary system are individual office holders, but it is important to distinguish the office from the individual who holds it. An office is, in principle, immortal; the holder is not. The individual in the course of his lifetime moves from office to office and always holds a variety of different "offices" simultaneously. The rights and duties of a particular office can be uniquely specified, but the rights and duties of a particular individual will depend on circumstance and will normally entail inconsistency. Thus an individual who holds office as head of a university department and is also the owner of a particular residential house and the father of a family will have rights and duties in respect to each of these statuses, but it may or may not be possible for him to fulfill all three roles with equal adequacy.

A full description of the social structure would entail an analysis of all the rights and obligations which link all the offices and corporations in the system—a task which would be plainly impossible but which can be carried out piecemeal in a partial way. The anthropologist's focus of interest is not in this structure as such but in "the way it works"—that is, in the perception of the way in which living human beings, who are all the time being born, growing older, and dying, *pass through* an ordered system of offices.

At a beginning level of analysis the system of offices is unchanged by the passage of individual office holders through the system, and it is at this level that Fortes has concentrated his own work (1949). At the next level of analysis "social change" results from the fact that individual office holders modify the rules pertaining to the offices which they occupy. In Fortes' view the sociological analysis of "change" in this sense is a matter for the future and may prove impossible.

Other British writers have tried to evade this



limitation, and recently there have been a number of attempts to bring the dimension of historical time within the scope of structural analysis while avoiding the pitfalls of historical determinism. Leach's *Political Systems of Highland Burma* (1954) was an attempt of this kind, which presents Kachin political ideology as a structure of persisting verbal categories and argues that historical change in the empirical society results from a re-interpretation of existing categories rather than from a change in the structure of ideas.

Fortes' interest in the way individuals move into and out of social office links his work with that of Gennep (1909), and, despite the greater importance attached to ideas deriving from Weber, the influence of early twentieth-century French sociology on the work of British "structuralist" anthropology remains very strong. As a consequence, despite differences of terminology and orientation, French and British ideas about social structure cannot be fully distinguished.

Lévi-Strauss and the French school. The mathematical idea of "structure," which is also that employed in "structural linguistics," is somewhat different from that of either architecture or biology. The emphasis here is upon the transformability of a relational set rather than upon the quality of the relations as such. If a piece of music—originally in the form of a sheet of printed notes—is played on a piano, recorded on a phonograph record, transmitted over the radio, and finally played back to an auditor, the music passes through a whole series of transformations—it appears as printed notes, as a pattern of finger movements, as sound waves, as modulations of the grooves on a piece of bakelite, as electromagnetic vibrations, and so on. However, something is common to all these manifestations. That common something is their structure. This mathematical idea is abstract and very general; it includes the biologist's notion of structure as a special case. The analogous notion of "social structure," when derived from the mathematical concept, has likewise more general, more abstract uses than when it is based on an organic analogy. Lévi-Strauss's view of structure has been taken over more or less directly from linguistics, and it is because of this that his arguments often appear highly abstract when compared to the relatively empirical analyses of British structural anthropologists, whose ideas ultimately derive from biology through the thinking of Radcliffe-Brown. Radcliffe-Brown conceived of the task of social anthropologists as that of establishing a kind of Linnaean taxonomy of functioning social systems. In such a taxonomy the difference between one

society and another would be one of structure, but this structure was a readily describable empirical characteristic of the system. In Lévi-Strauss's analysis, on the other hand, interest in structure comes to the fore only when one system is contrasted with another. Lévi-Strauss is interested in how far empirical systems correspond (or fail to correspond) with the theoretically possible transformations of a single (mathematical) structure; his structures thus start out as "models" rather than as empirically observed facts. For example, it is possible to imagine, as an ideal type, a society which has a strictly unilineal descent system and a strictly "unilocal" residence pattern. This gives us two binary variables—descent, which is patrilineal ( $P$ ) or matrilineal ( $M$ ), and residence, which is virilocal ( $V$ ) or uxorilocal ( $U$ ). With these limitations, we have four possibilities (or classes):  $P + V$ ;  $M + U$ ;  $P + U$ ;  $M + V$ . Actual societies do not fit tidily into categories of this kind, yet a great many actual systems can be approximately sorted in this way. If this is done certain facts emerge. The distribution is not at all random. Societies of class 1 are common, of class 4 less common, of class 2 very rare, and of class 3 virtually nonexistent.

We are then led to consider just why such distributions should be nonrandom, and this may lead us to fundamental insights into factors which affect all human societies and not just societies of a particular taxonomic type.

In a Radcliffe-Brown analysis, "the structure which persists" is the structure of a particular society associated with a particular culture and a particular geographical locality; in a Lévi-Strauss analysis, "the structure which persists" is an attribute of human social organization as such; we can come to recognize the nature of this structure only when we build up a generalized model from radically contrasted empirical examples. As a case in point, Lévi-Strauss (1963) has compared the stereotype of Australian aboriginal society with the stereotype of Indian caste (see Table 1).

Lévi-Strauss's argument is that in the Australian context all the items in the left-hand column are structurally and functionally interdependent and in the Indian context all the items in the right-hand column are similarly interdependent, but since the items listed on the right are the mirror inversion of those listed on the left, the structure represented by the right-hand column is really the same structure as that represented by the left-hand column. The right-hand column is generated from the left-hand column by applying a single transformation rule—the structure "remains the same," in the same

Table 1

Australian society	Indian society
In Australia the social unit is a "horde," which is exogamous.	In India the social unit is a "sub-caste" ( <i>jati</i> ), which is endogamous.
Hordes are distinguished as totemic groups, totems being categories of nature (animals, plants).	Subcastes are distinguished as occupational groups, occupations being categories of culture (potters, ropemakers, drummers, etc.).
The activities of totemic groups are group activities sanctioned by myth and consist of the ritual creation of animals (things of nature).	The activities of caste groups are individual activities based on real knowledge and consist of the rational creation of artifacts (things of culture).
Magical totemic increase ceremonies rest on an ideology that totemic groups should supply each other with food just as they supply each other with women.	Ritual attitudes to pollution have the consequence that a subcaste is a commensal group as well as an endogamous group. The exchange of food across subcaste boundaries is prohibited.
In all other respects the division of labor depends only upon sex and age. Horde groups and totemic groups are not occupationally differentiated.	Since subcastes are differentiated by occupational specialization they are economically interdependent.
Thus the larger society is united by intergroup transactions involving the interchange of objects of nature (women and food).	Thus the larger society is united by intergroup transactions involving the interchange of objects of culture (artifacts and occupational services).
Horde groups and totemic groups are of intrinsically equal status.	Subcastes are of intrinsically unequal status.

Source: Based on Lévi-Strauss 1963.

sense as a square remains the same when we rotate it through 90 degrees or an algebraic equation remains the same when we change all plus signs into minus and minus signs into plus. Lévi-Strauss's procedures for discriminating between relevant and irrelevant evidence are no doubt arbitrary, but so are those of his British colleagues, and his demonstration that patterns of "structure" of this kind exist in the ethnographic data poses problems of fundamental sociological significance.

**Assessment.** I have outlined three different uses to which the concept of social structure has been put by leading contemporary anthropologists. The value which we place on these different usages will depend upon our objectives. The Murdock usage and the Radcliffe-Brown usage are clearly to be preferred if our prime interest is typological and taxonomic, while those who set themselves the ambitious objective of discovering human universals will be more attracted by the higher-level abstractions which Lévi-Strauss has brought into

the discussion. It deserves note, however, that none of these theorists has as yet dealt at all adequately with the time dimension in social affairs.

Whichever way we approach the concept of structure, it entails the notion of persistence. In Fortes' treatment, the time of history and the time of the individual life cycle are distinguished. The structure of social units is conceived of as permanent relative to the life span of the individual members of society who pass through the system. Fortes would not claim that social structures in his sense are changeless but only that they are relatively changeless. For his purposes, change in structure is a phenomenon of history and as such lies outside sociological analysis. Lévi-Strauss, on the other hand, has tried to integrate his scheme with a kind of Neo-Marxist-Hegelian dialectic. The emphasis on binary contrast, which permeates the "structuralism" of Lévi-Strauss and his associates in a very radical way, is a direct borrowing from the structural linguistics of the Prague school and is closely linked with Jakobson's use of the notion of "distinctive features." It also seems to tie in with the *contrast theory of meaning*, which is one of the doctrines of the Oxford school of linguistic philosophy. The essence of the argument is, roughly, that "we can recognize a thing for what it is by seeing what it is not"; thus the structure of any set is that which is common both to the set and its inverse (cf. the horde/subcaste example above). This can be given a Hegelian twist by identifying the structure with the "synthesis" which derives from (but is already implicit in) any "thesis" when viewed in conjunction with the "antithesis" which it generates. It is not at all clear to the present writer just how Lévi-Strauss relates this potentially historicist argument to his structural models. Sometimes he seems to be saying simply that contrasted empirical examples of social systems are best understood as different manifestations of a single model structure; at other times there is a suggestion that the different manifestations are viewed as a dialectical sequence, each empirical type serving to generate its own antithesis. By this argument societies which are contrasted manifestations of the same structure should be found distributed in adjacent territorial areas in a synchronic dimension and in the same territorial area in a diachronic dimension. Pareto's ideas about a "moving equilibrium" are not very different from this. Leach's *Political Systems of Highland Burma* (1954) can be considered as an attempt to test out this argument in terms of concrete ethnohistorical materials.

Structuralism in linguistics has run into the same time-scale difficulties as structuralism in



social anthropology, and it may be of interest to see what has happened in this neighboring field. It has lately come to be emphasized that a verbal utterance can never be a synchronic event. It starts at the beginning and progresses to the end. The form of the utterance is not, in general, completely predetermined by its beginning; it is "generated" as the utterance progresses as a response to grammatical rules which are applicable or not applicable depending upon the developing sequence. The grammatical rules are like the rules of a game: they do not determine what is said; they merely delimit what is a permissible choice at any particular point in the sequence. They specify what is possible and also what is not possible, but between these limits they allow freedom of choice. The work of the mathematical economists has been developing in a rather similar way: progressions of events through time are coming to be thought of as delimited by qualified rules (as in a game) rather than "determined" in an historicist sense. This kind of application of game theory and its derivatives disposes of the notion that structure is, of necessity, a static synchronic concept. A game of chess is a sequence in time, and no individual move in the sequence is predictable, yet it is plain that the sequence as a whole is "structured" in a quite formal and easily describable way. The rules of chess are specific and not generative, but we are familiar enough with types of games where the rules themselves change according to circumstance—e.g., a dice game which includes the rule "player who throws a six has two extra throws." Some British anthropologists are already disposed to make their descriptions of social structure in terms of sets of rules defining the interdependent rights and obligations attaching to sets of offices. To start thinking of such rules as being of the generative kind would not call for any very drastic modification of established conventions. Barth (1959; 1966) is one of the few anthropologists who has so far made some tentative exploration of this idea. In the view of the present writer it may prove a fruitful line of development.

EDMUND R. LEACH

#### BIBLIOGRAPHY

- BARTH, FREDRIK 1959 Segmentary Opposition and the Theory of Games: A Study of Pathan Organization. *Journal of the Royal Anthropological Institute of Great Britain and Ireland* 89:5-21.
- BARTH, FREDRIK 1966 *Models of Social Organization*. Royal Anthropological Institute of Great Britain and Ireland, Occasional Papers, No. 23. London: The Institute.
- BATESON, GREGORY (1936) 1958 *Naven: A Survey of the Problems Suggested by a Composite Picture of the Culture of a New Guinea Tribe Drawn From Three Points of View*. 2d ed. Stanford Univ. Press.
- DAVENPORT, WILLIAM 1959 Nonunilinear Descent and Descent Groups. *American Anthropologist* New Series 61:557-572.
- ENGELS, FRIEDRICH (1884) 1940 *The Origin of the Family, Private Property and the State*. London: Lawrence & Wishart.
- EVANS-PRITCHARD, E. E. (1940) 1963 *The Nuer: A Description of the Modes of Livelihood and Political Institutions of a Nilotic People*. Oxford: Clarendon.
- FALLERS, LLOYD A. 1956 *Bantu Bureaucracy: A Study of Integration and Conflict in the Political Institutions of an East African People*. Cambridge: Heffer.
- FIRTH, RAYMOND 1964 *Essays on Social Organization and Values*. London School of Economics and Political Science, Monographs on Social Anthropology, No. 28. London: Athlone.
- FORDE, DARYLL 1950 Double Descent Among the Yakö. Pages 285-332 in A. R. Radcliffe-Brown and Daryll Forde (editors), *African Systems of Kinship and Marriage*. Oxford Univ. Press.
- FORDE, DARYLL 1963 Unilinear Fact or Fiction: An Analysis of the Composition of Kin-groups Among the Yakö. Pages 38-57 in Isaac Schapera (editor), *Studies in Kinship and Marriage*. London: Royal Anthropological Institute of Great Britain and Ireland.
- FORTES, MEYER 1945 *The Dynamics of Clanship Among the Tallensi*. Oxford Univ. Press.
- FORTES, MEYER (1949) 1963 Time and Social Structure: An Ashanti Case Study. Pages 54-84 in Meyer Fortes (editor), *Social Structure: Studies Presented to A. R. Radcliffe-Brown*. New York: Russell.
- FORTES, MEYER 1953 The Structure of Unilinear Descent Groups. *American Anthropologist* New Series 55:17-41.
- FORTES, MEYER; and EVANS-PRITCHARD, E. E. (editors) (1940) 1958 *African Political Systems*. Oxford Univ. Press.
- GENNEP, ARNOLD VAN (1909) 1960 *The Rites of Passage*. Univ. of Chicago Press; London: Routledge. → First published in French.
- GOODY, JACK R. (editor) 1958 *The Developmental Cycle in Domestic Groups*. Cambridge Papers in Social Anthropology, No. 1. Cambridge Univ. Press.
- LEACH, EDMUND R. 1954 *Political Systems of Highland Burma: A Study of Kachin Social Structure*. Cambridge, Mass.: Harvard Univ. Press; London School of Economics and Political Science.
- LEFEBVRE, H. 1962 Le concept de structure chez Marx. Pages 100-106 in Colloque sur le mot structure, Paris, 1959, *Sens et usages du terme structure dans les sciences humaines et sociales*. The Hague: Mouton.
- LÉVI-STRAUSS, CLAUDE (1958) 1963 *Structural Anthropology*. New York: Basic Books. → First published in French.
- LÉVI-STRAUSS, CLAUDE 1963 The Bear and the Barber. *Journal of the Royal Anthropological Institute of Great Britain and Ireland* 93:1-11.
- LINTON, RALPH 1936 *The Study of Man: An Introduction*. New York: Appleton.
- MAINE, HENRY J. S. (1861) 1960 *Ancient Law: Its Connection With the Early History of Society, and Its Relations to Modern Ideas*. Rev. ed. New York: Dut-

- ton; London and Toronto: Dent. → A paperback edition was published in 1963 by Beacon.
- MARX, KARL (1859) 1913 *A Contribution to the Critique of Political Economy*. 2d ed. Chicago: Kerr. → See especially Marx's "Preface" in the 1913 edition.
- MORGAN, LEWIS H. 1871 *Systems of Consanguinity and Affinity of the Human Family*. Smithsonian Institution Contributions to Knowledge, Vol. 17; Publication No. 218. Washington: The Institution.
- MORGAN, LEWIS H. (1877) 1963 *Ancient Society*. 2d ed. New York: Meridian.
- MURDOCK, GEORGE P. 1949 *Social Structure*. New York: Macmillan. → A paperback edition was published in 1965 by the Free Press.
- MURDOCK, GEORGE P. 1955 Changing Emphasis in Social Structure. *Southwestern Journal of Anthropology* 11:361-370.
- MURDOCK, GEORGE P. (editor) 1960 *Social Structure in Southeast Asia*. Viking Fund Publications in Anthropology, No. 29. Chicago: Quadrangle Books.
- NADEL, S. F. 1957 *The Theory of Social Structure*. London: Cohen & West; Glencoe, Ill.: Free Press.
- RADCLIFFE-BROWN, A. R. (1924-1949) 1952 *Structure and Function in Primitive Society: Essays and Addresses*. London: Cohen & West; Glencoe, Ill.: Free Press.
- RADCLIFFE-BROWN, A. R. 1957 *A Natural Science of Society*. Glencoe, Ill.: Free Press.
- RADCLIFFE-BROWN, A. R.; and FORDE, DARYLL (editors) 1950 *African Systems of Kinship and Marriage*. Published for the International African Institute. Oxford Univ. Press.
- SPENCER, HERBERT (1858) 1966 *Prospectus of a System of Philosophy Issued by Herbert Spencer in 1858*. Pages 297-303 in Jay Rumney, *Herbert Spencer's Sociology: A Study in the History of Social Theory*. New York: Atherton.
- VOGT, EVON Z. 1960 On the Concept of Structure and Process in Cultural Anthropology. *American Anthropologist* New Series 62:18-33.
- WARNER, W. LLOYD (1937) 1958 *A Black Civilization: A Social Study of an Australian Tribe*. Rev. ed. New York: Harper.
- YALE UNIVERSITY, INSTITUTE OF HUMAN RELATIONS (1938) 1961 *Outline of Cultural Materials*. 4th rev. ed. New Haven: Human Relations Area Files Press.

## II

### SOCIAL STRUCTURAL ANALYSIS

The concept "social structure" is, paradoxically, so fundamental to social science as to render its uncontested definition virtually impossible. Basic concepts in particular seem likely to suffer from this difficulty, since their primordial character demands that they provide an effective link between the field of inquiry and the particular approach, or even the personal philosophy, of the individual scholar. Various controversies, both apparent and real, are thus especially likely to surround the use of such concepts. In the case of "social structure," some of these difficulties are purely terminological and are, hence, more easily resolved than may

appear possible on the surface. Others derive more fundamentally from disagreement over the basic philosophical assumptions appropriate to social research. And still others reflect substantive theoretical problems, which seem likely to persist even when there is agreement over terminology and basic philosophy. Before one can proceed to analyze social structure in detail, one must confront certain basic issues arising from these three areas of controversy.

**Terminology.** Despite some contention to the contrary, I would hold that terminological problems are relatively easy to solve, since there is more agreement among sociologists and other behavioral scientists about concepts than about what words should be used to designate them. Virtually every general approach to social behavior involves some or all of the following five conceptual distinctions as actual or potential distinct loci of variation:

- (1) An *individual* component, i.e., some conception of personality or aggregated characteristics of individual personalities.
- (2) A *group* component, i.e., a set of variables essentially referring to various aspects of the process of social interaction among people.
- (3) A *morphological* component, i.e., a set of variables pertaining to the spatial-temporal arrangement of individuals and the physical size of groups.
- (4) A *systemic* component which focuses on the properties of interrelationships among activities as such, independently of the people performing them, generally treating such activities as social roles.
- (5) A *cultural* component comprising ideas (norms, values, beliefs) which are learned and shared by people and are transmitted symbolically from person to person.

As we shall see presently, different philosophical and theoretical viewpoints yield different orderings of relative emphasis on these five components and, indeed, sometimes result in wholly eliminating one or another of the components from effective consideration. From a purely definitional standpoint, the question of what one means by "social structure" thus reduces to the question of how many and which of these five conceptual components one wishes to denote by the term. Since we wish to compare, and to some degree synthesize, various points of view, we shall define "social structure" very broadly as the totality of patterns of collective human phenomena that cannot be explained solely on the basis of human heredity and/or the non-human environment (compare Levy 1952, pp. 8-17). This definition includes all five aspects, allows



for interrelations among them, and thus provides a context for a comparative discussion of current approaches.

### Philosophical assumptions

Certain basic philosophical questions raised principally by the "structuralist" cultural anthropologists bear on the question of the relationship of models to social structure (Lévi-Strauss [1958] 1963, pp. 277-323). Everyone presumably agrees that the twentieth-century scientist, as opposed to his nineteenth-century counterpart, is no longer trying to "discover reality" but, rather, is seeking to understand observations by imposing different kinds of order on them in the form of various models, exploring the implications of each, and accepting one model as opposed to another on pragmatic grounds. The structuralist position in effect extends this argument by observing that ordinary people are subject to the same epistemological limitations as are scientists. According to this view, any discussion or thought of social structure by anyone implies that he is constructing and using models, which we shall term "folk models" to distinguish them from the analytic models constructed by the scientist (Anderson & Moore 1960). Folk models thus constitute an important cultural element of social structure; some examples would be forms of games and patterns of responses to questions, as well as language itself. The formal properties of folk models are, furthermore, susceptible to study and are capable of yielding important insights into social structure through such procedures as, for example, "componential analysis" (Goodenough 1956).

Few social scientists would disagree with this position as stated so far. To be sure, it begs the question of the relative explanatory importance of folk models as opposed to other aspects of social structure under given circumstances, but such a question is presumably a fruitful one which can in principle be answered through empirical research. However, the argument usually does not stop at this point but frequently proceeds to combine with the foregoing an essentially phenomenological view of social reality. As a result, folk models emerge, not simply as one important component of social structure but as the totality of social structure itself. The study of social structure thus becomes no more and no less than the study of folk models (see Lévi-Strauss [1958] 1963, p. 279). This perspective leaves us, however, with the problem of the relationship between folk models, on the one hand, and analytic models constructed by the scientific observer, on the other.

It must be said that some confusion occurs in the structuralist literature as a result of a frequent failure to make this distinction explicit. Once such a distinction is made, however, it becomes apparent that these two types of models are very likely to be quite different; indeed, if they were not, the value of social science itself would be highly questionable. And not only can the social scientist construct analytic models based on his analysis of folk models, but he can also construct analytic models of other social phenomena that are not necessarily reflected in the cultural system at all, such as certain morphological patterns. In order to maintain a consistently phenomenological viewpoint in the face of such possibilities, the structuralists are obliged at this point to assume that scientific models have latent cultural existence as "unconscious" folk models (*ibid.*, p. 281).

It is certainly possible, in principle, to proceed in this manner, but one may reasonably question the value of so doing. If one is concerned primarily with the analysis of the formal properties of symbolic systems *per se*—as are, in all fairness, most structuralists—the difficulties of this mode of procedure are not great. But where the analysis requires any investigation of latent behavioral or morphological characteristics, the attempt to maintain a consistently phenomenological viewpoint would seem to involve many unnecessary complications. Pragmatically, the same results could be obtained more easily by merely assuming such behavioral and morphological characteristics to be external to the cultural system and raising purely empirical questions regarding their relationships to it. Furthermore, divesting the structuralist approach of phenomenology does not result in stripping it of its genuinely valuable contributions: that it calls attention to the inevitability of model construction in the study of social structure and that it identifies folk models as an important component of cultural systems, in the process drawing particular attention to the relevance of linguistic analysis (Anderson & Moore 1960; Dumazedier 1964; Jaeger & Selznick 1964; Osgood et al. 1957). Moreover, such divestiture reveals a very important question that otherwise remains confused or obscured—the question of the importance of folk models, relative to other patterns of social phenomena, as sources of theoretical explanation. The problem of the relationship, under various conditions, of folk models on the cultural level to other patterns manifest on the morphological, systemic, group, and individual levels of social structure is a matter for empirical investigation, the answer to which can hardly be assumed away or defined out

of existence. Thus I would accept the epistemological observations of structuralism but would stop short of a phenomenological position. This position leads to substantive agreement with the structuralists that folk models, conceived as part of the cultural component of social structure, are indeed important; however, it also raises the question of the relative importance of such models under specified conditions as a matter for empirical determination.

### Theoretical problems

We may now turn to a discussion of actual theoretical issues as they center on two major questions: first, the relative ordering of, and general relationships among, the five aspects of structure previously indicated; and, second, the nature of the mechanisms of interrelationship.

**Divergent views of social structure.** As to the questions of ordering and general relationships, three basic viewpoints have been particularly important; I shall refer to them as the cultural, interactionist, and morphological positions, respectively. The *cultural* point of view emphasizes the explanatory primacy of culture as a system of ideas that are learned, shared, and transmitted by people who, in turn, are oriented to situations by virtue of the culture that they have learned. Situations thus evoke responses, culturally defined as role expectations, by virtue of which people find themselves interacting with one another, with the possibility of resultant morphological differentiation. The causal progression is from the cultural-individual aspects of the situation to the systemic aspects, to group characteristics, and, finally, to the morphological aspects of structure. The entire chain is not usually explicit; it is more common for this approach to stress the interrelationships of culture, personality, and social system, defining social structure as some combination of these elements, with group and morphological considerations viewed as somewhat ancillary. An essentially "theatrical" view of social structure emerges, wherein culturally defined systems of role expectations await the participation of people, and this participation either incidentally yields group and morphological characteristics or implies them as contextual constants.

It is impossible to discuss here all of the divergences of detail exhibited by different versions of the cultural approach; major instances would include "culturological" social anthropology (e.g., Kluckhohn 1953; Kluckhohn & Strodtbeck 1961; Murdock 1949), the phenomenological structuralism previously discussed, and the voluntaristic

theory of action applied to the analysis of social structure (Weber 1922; Parsons 1951; for a similar but somewhat different approach, see Nadel 1957; also see Kroeber & Parsons 1958).

In some respects both the interactionist and the morphological positions represent the obverse of the cultural approach, in that both postures either imply cultural and individual characteristics as contextual constants or view them as emergent secondary properties. In the *interactionist* version of social structure, people with highly generalized expectations about one another's probable behavior are taken as given and are morphologically presumed to interact with one another. From the ensuing reciprocal process of role taking and role playing, increasingly stable systems of role expectations emerge which, in turn, become generalized into superordinate systems of cultural norms. The causal progression is thus from group phenomena to systemic aspects and, finally, to cultural patterns (Blau 1964; Homans 1961; Mead 1934).

The *morphological* conception of social structure is essentially similar to the interactionist version, but it is more macroscopically conceived. It assumes a disposition on the part of people to engage in stable interaction relative to certain specified but highly generalized cultural expectations. Emergent systemic and cultural properties of the situation therefore depend on morphological structure, which, through variations in the numbers of people involved and their spatial and temporal relationships to one another, as well as to artifactual facilities, imposes limits on the possibilities of interaction capable of producing emergent characteristics (Cottrell 1955; Duncan & Schnore 1959; Durkheim 1893; Halbwachs 1938; Helm 1962; Schnore 1958).

These divergent viewpoints have, in turn, yielded apparently opposing conceptions of social structure. From the cultural standpoint, social structure emerges as an object of orientation comprising cultural and systemic elements in terms of which the actions of individuals, groups, and morphological entities occur or are directed. On the other hand, from the interactionist and morphological viewpoints, social structure has, by and large, been thought of as the totality of group and morphological patterns, with individual elements contained in the group patterns and with or without systemic components, depending on the particular version of theory. Cultural elements are viewed as essentially external, in either a contextual or an emergent sense, or they are for some purposes deemed irrelevant.

Two principal theoretical controversies have re-



sulted from this situation; one of these may be regarded as at least partially resolved, but the other is now only beginning to be fully explored. The first of these controversies is that of the cultural versus the interactionist versions of social structure. The key to the resolution of this problem lies in the development of the voluntaristic theory of action combined with the introduction of symbols into interaction theory. No adequate theory of social behavior now seems possible unless it explicitly recognizes some process of social learning. The symbolic interactionist position, in contrast with that at least implied by earlier "radical positivists," is that social interaction cannot be a purely physical process but, rather, involves reciprocal role taking and role playing with reference to a learned repertoire of relatively stable expectations (Mead 1934).

Approaching the same problem from the opposite pole, the voluntaristic theory of action seeks to remove the metaphysical cobwebs from transcendental idealism by arguing that the "force" of ideas on the social order is actually felt in an entirely empirical manner—through socialization and subsequent action, with inevitable reference to learned normative orientations (Parsons 1937, esp. pp. 77–82). Both positions, therefore, not only insist upon the existence of cultural components in social structure but also identify essentially similar ones: relatively stable sets of learned (and "learnable") normative expectations. The result is a more comprehensive view of social structure than either the cultural or the interactionist position alone provides; it includes four of the five components we have identified, excluding only morphological structure.

It is precisely this absence of morphological considerations that leads to the second major theoretical issue: the place of human ecology in current theories of social structure. The emphasis on elements capable of transmission through learning has at times led to stressing roles and norms to the exclusion of concrete people and things in physical space. The result can be a view of social action that is literally carried on by nobody anywhere. In this sense the achievement of a measure of integration between cultural and interactionist viewpoints has been something of a Pyrrhic victory; it has concomitantly involved—at least on the theoretical level—a gradual destruction of an earlier positivist unity of morphological and interactionist approaches (Durkheim 1893).

Among the consequences of this situation has been a tendency either to view demography and human ecology as distinct disciplines entirely sep-

arate from sociology, or to view many current theoretical developments as essentially irrelevant to social reality. Furthermore, since social research must necessarily start by observing people and things, another consequence of this line of theoretical development has at times been a virtually enforced lack of relationship between research and theory. This situation has led unavoidably to the existence of entire bodies of analytical and statistical operations that have no clear correlates in current theoretical thought and, consequently, to the viewpoint—which, in light of the highly operationalistic emphasis of modern science, is surprisingly widely accepted in sociology—that there exists something called "methodology" which is generically different from something else called "theory." The unreality of this dichotomy has made increasingly apparent the fact that the cultural, interactionist, and morphological approaches do not actually represent differing "schools of thought" so much as they represent emphases on different aspects of social structure; thus, any reasonably complete theory of social structure must explicitly confront the question of interrelationships among these different aspects.

**Causal mechanisms.** Curiously enough, the question of causal mechanisms in social structure has received relatively little explicit attention. Broadly speaking, two kinds of mechanisms seem to pervade the literature, more often implicitly than explicitly, deriving from the cultural and morphological approaches, respectively. The first type of mechanism, which we shall term *psychological*, assumes that individuals severally receive cues from the social environment which, when combined in their personalities with learned responses (culturally prescribed or not), result in actions that, when aggregated, constitute social-structural patterns. In contrast, the second type of mechanism, which we shall term *ecological*, seeks to explain behavior through a process of elimination by arguing that certain actions are impossible owing to limiting features of the situation; in other words, social action is patterned by virtue of the fact that it must occur within the confines of limiting conditions.

To some extent it is true that cultural theories of social structure tend to rest on psychological explanatory mechanisms and morphological theories to rest on ecological mechanisms, whereas interactionist theories historically have exhibited a shift in emphasis from ecological to psychological mechanisms. But this statement is, at best, only a partial approximation. For example, ecological mechanisms need not be limited to those involving

physical properties of a morphological nature; they can be extended to include cultural and systemic characteristics, as in "structural-functional requisite analysis" (Levy 1952) and in adaptive theories of social change. Wilson and Wilson, for instance, explicitly define "social structure" as "the systematic form of limitation by which eccentricities are checked and complementary diversities are preserved" ([1945] 1965, p. 49, italics added; also see Eisenstadt 1964). Actually, it is impossible to discuss one type of mechanism meaningfully without at least making some implicit assumptions about the way the other is operating in the situation. Thus the scope of any psychological mechanism is limited by what cues are in fact present in the situation and what past elements could have been learned, as well as by physical, including artifactual, possibilities. Similarly, all ecological mechanisms assume the presence of some given level and kind of motivational orientation on the part of the actors involved. Thus the main theoretical issue is not which mechanism is present in some absolute sense but, rather, at what point in the analysis it is pragmatically appropriate to explicate which type of mechanism while making what assumptions about the other type.

This issue has often been obscured, perhaps partly because of a lack of synthetic effort, but surely because of the intellectual legacy of functionalism. For in its classical statement functionalism draws attention away from the question of causal mechanisms simply by assuming that social structures operate in such a way as to maintain themselves in some stable state (Merton [1949] 1957, pp. 19-84; Radcliffe-Brown 1952). Such an assumption permits one to make the type of ecological argument characteristic of "structural-functional requisite analysis": that the existence of such a state implies that certain activities contributing to its maintenance are perforce somehow being performed (Levy 1952, pp. 149-197). But such an argument invites questions about the status of psychological mechanisms in the situation (Brown 1963, pp. 109-132; Homans 1964; Inkeles 1959). Classical functionalism avoided such questions by assuming, in effect, a complete congruence of psychological mechanisms with ecological explanatory devices, thereby giving a teleological cast to social structure (Malinowski 1944) and inviting tautologous formulations. It is possible to drop this assumption and still maintain a nominally functionalist position. But the questions begged as a result of so doing transcend the original functionalist position to such a degree as to suggest the adoption of a different perspective al-

together. The focus becomes one of discovering the nature of causal mechanisms that interconnect different aspects of social structure, rather than one of assuming the mechanisms and thereby inferring structures (Dore 1961).

This position suggests a resolution, in principle, of another question in regard to causal mechanisms that has often plagued the analysis of social structure: the "statics versus dynamics" problem. Masking the theoretical issues is the immediate terminological question of whether the concept "social structure" should be restricted in its reference only to those elements of the social order which remain constant over time. Thus Firth (1951), for example, proposes to use the term "social organization" to denote patterns of change, in contrast with "social structure," which indicates patterns of stable elements. Analogous to this distinction is the familiar "process versus structure" dichotomy. But processes are inseparable from structure, in the sense that any social structure is held together by mechanisms relating its elements (see Vogt 1960; see also Gurwitsch [1950] 1957-1963, vol. 2, pp. 403-446). Operation of such mechanisms through time results in more or less change or stability in the values of the variables reflecting the social structure. Consequently it is misleading to treat patterns of stable interrelations as if they were basically different from patterns of change, inasmuch as both types of pattern result from the operation of social forces through causal mechanisms. It is not the case that change, simply because it is change, somehow involves "more mechanisms" than does stability.

### Toward an eclectic view

The current path of what one might term "post-functional revisionism" thus suggests that dynamic strain is always present on some levels of social structure (Moore 1963), and, while hardly Marxist in its orientation, suggests the fruitfulness of a reconsideration of some of the basic insights of Marx in this context. The five structural aspects which we have distinguished are interrelated by a complex of dialectical tensions which, in principle, can be concretely operationalized and explicated as a set of interrelated psychological and ecological causal chains cutting across the five aspects of social structure, with their operation resulting in more or less "stability" and "change."

The immediate problem would seem to be, first, the allocation of proportional contributions of variance among the cultural, systemic, morphological, group, and individual aspects of social structure. One can then seek to explain the statistical results



thus obtained by assuming various psychological and ecological causal mechanisms whose existence can subsequently be verified or disproved by detailed analysis of actual social situations.

An as yet small, but perhaps significant, literature is beginning to arise utilizing essentially this perspective and centering—at least in its initial stages—on the analysis of “structural” or “compositional” effects. Thus far it has proved possible, in such analyses, to isolate individual from group effects and to separate both individual and group effects from a residue involving morphological, systemic, and cultural properties (Blau 1960; Lazarsfeld & Menzel 1961; Davis et al. 1961). Some inferences of precise mechanisms have been made, mostly centering on socialization on the individual level; interpersonal strategies yielding “social approval” on the group level; and emergent consolidation of normative systems through adaptation between the group and systemic, and group and morphological, levels (Davis 1963; also see Heider 1958; Blau 1964).

It is as yet too early to assess the probable ultimate fruitfulness of this approach. Various problems seem to arise, for example, in developing consistent operational ways of moving from the group to the systemic, morphological, and cultural levels of analysis. The main difficulty seems to lie in systematically relating interpersonal phenomena to ecological processes (in the traditional sense), value systems, and various (often vaguely conceived) macrostructural, dialectical “forces” (Boudon 1963; Selvin & Hagstrom 1963; Tannenbaum & Bachman 1964). However, with increasing recognition that many theoretical controversies regarding social structure either reduce to the use of different words to denote the same things or arise from an emphasis on different relatable aspects of the same phenomena, will perhaps come a fruitful synthesis, eventually more original than eclectic.

STANLEY H. UDY, JR.

[See also FUNCTIONAL ANALYSIS; INTERACTION, article on SOCIAL INTERACTION; PERCEPTION, article on SOCIAL PERCEPTION; SOCIAL INSTITUTIONS; SOCIOLOGY, article on THE DEVELOPMENT OF SOCIOLOGICAL THOUGHT.]

#### BIBLIOGRAPHY

- ANDERSON, ALLAN R.; and MOORE, OMAR K. 1960 Autotelic Folk-models. *Sociological Quarterly* 1:203-216.
- BLAU, PETER M. 1960 Structural Effects. *American Sociological Review* 25:178-193.
- BLAU, PETER M. 1964 *Exchange and Power in Social Life*. New York: Wiley.
- BOUDON, RAYMOND 1963 Propriétés individuelles et propriétés collectives: Un problème d'analyse écologique. *Revue française de sociologie* 4:275-299.
- BROWN, ROBERT R. 1963 *Explanation in Social Science*. Chicago: Aldine.
- COTTRELL, WILLIAM F. 1955 *Energy and Society*. New York: McGraw-Hill.
- DAVIS, JAMES A. 1963 Structural Balance, Mechanical Solidarity, and Interpersonal Relations. *American Journal of Sociology* 68:444-462.
- DAVIS, JAMES A.; SPAETER, JOE L.; and HUSON, CAROLYN 1961 A Technique for Analyzing the Effects of Group Composition. *American Sociological Review* 26:215-225.
- DORE, RONALD P. 1961 Function and Cause. *American Sociological Review* 26:843-853.
- DUMAZEDIER, JOFFRE 1964 Structures lexicales et significations complexes. *Revue française de sociologie* 5:12-26.
- DUNCAN, OTIS DUDLEY (1964) 1966 Social Organization and the Ecosystem. Pages 36-82 in Robert E. L. Faris (editor), *Handbook of Modern Sociology*. Chicago: Rand McNally.
- DUNCAN, OTIS DUDLEY; and SCHNORE, LEO F. 1959 Cultural, Behavioral, and Ecological Perspectives in the Study of Social Organization. *American Journal of Sociology* 65:132-146. → A comment by Peter H. Ross and a rejoinder by Otis Dudley Duncan and Leo F. Schnore appear on pages 146-153.
- DURKHEIM, ÉMILE (1893) 1960 *The Division of Labor in Society*. Glencoe, Ill.: Free Press. → First published as *De la division du travail social*.
- EISENSTADT, SHMUEL N. 1964 Social Change, Differentiation, and Evolution. *American Sociological Review* 29:375-386.
- FIRTH, RAYMOND W. 1951 *Elements of Social Organization*. London: Watts. → A paperback edition was published in 1963 by Beacon.
- GOODENOUGH, WARD H. 1956 Componential Analysis and the Study of Meaning. *Language* 32:195-216.
- GURVITCH, GEORGES (1950) 1957-1963 *La vocation actuelle de la sociologie*. 2d ed. 2 vols. Paris: Presses Universitaires de France. → Volume 1: *Vers la sociologie différentielle*. Volume 2: *Antécédents et perspectives*.
- HALEWACHS, MAURICE (1938) 1960 *Population and Society*. Glencoe, Ill.: Free Press. → First published as *Morphologie sociale*.
- HEIDER, FRITZ 1958 *The Psychology of Interpersonal Relations*. New York: Wiley.
- HELM, JUNE 1962 The Ecological Approach in Anthropology. *American Journal of Sociology* 67:630-639.
- HOMANS, GEORGE C. 1961 *Social Behavior: Its Elementary Forms*. New York: Harcourt.
- HOMANS, GEORGE C. 1964 Bringing Men Back In. *American Sociological Review* 29:809-818.
- INKELES, ALEX 1959 Personality and Social Structure. Pages 249-276 in American Sociological Society, *Sociology Today: Problems and Prospects*. Edited by Robert K. Merton, Leonard Broome, and Leonard S. Cottrell, Jr. New York: Basic Books.
- JAEGER, GERTRUDE; and SELZNICK, PHILIP 1964 A Normative Theory of Culture. *American Sociological Review* 29:653-669.
- KLUCKHOHN, CLYDE (1953) 1961 Universal Categories of Culture. Pages 89-105 in Frank W. Moore (editor), *Readings in Cross-cultural Methodology*. New Haven: Human Relations Area Files Press.
- KLUCKHOHN, FLORENCE R.; and STRODTBECK, FRED L. 1961 *Variations in Value Orientations*. Evanston, Ill.: Row, Peterson.

- KROEBER, A. L.; and PARSONS, TALCOTT 1958 *The Concepts of Culture and of Social System*. *American Sociological Review* 23:582-583.
- LAZARSFELD, PAUL F.; and MENZEL, HERBERT 1961 *On the Relation Between Individual and Collective Properties*. Pages 422-440 in Amitai Etzioni (editor), *Complex Organizations: A Sociological Reader*. New York: Holt.
- LÉVI-STRAUSS, CLAUDE (1958) 1963 *Structural Anthropology*. New York: Basic Books. → First published in French.
- LEVY, MARION J. JR. 1952 *The Structure of Society*. Princeton Univ. Press.
- MALINOWSKI, BRONISLAW 1944 *A Scientific Theory of Culture and Other Essays*. Chapel Hill: Univ. of North Carolina Press.
- MEAD, GEORGE H. (1934) 1963 *Mind, Self and Society From the Standpoint of a Social Behaviorist*. Edited by Charles W. Morris. Univ. of Chicago Press. → Published posthumously.
- MERTON, ROBERT K. (1949) 1957 *Social Theory and Social Structure*. Rev. & enl. ed. Glencoe, Ill.: Free Press.
- MOORE, WILBERT E. 1963 *Social Change*. Englewood Cliffs, N.J.: Prentice-Hall.
- MURDOCK, GEORGE P. 1949 *Social Structure*. New York: Macmillan. → A paperback edition was published in 1965 by the Free Press.
- NADEL, S. F. 1957 *The Theory of Social Structure*. London: Cohen & West; Glencoe, Ill.: Free Press. → Published posthumously.
- OSGOOD, CHARLES E.; SUCI, G. J.; and TANNENBAUM, P. H. (1957) 1961 *The Measurement of Meaning*. Urbana: Univ. of Illinois Press.
- PARSONS, TALCOTT (1937) 1949 *The Structure of Social Action: A Study in Social Theory With Special Reference to a Group of Recent European Writers*. Glencoe, Ill.: Free Press.
- PARSONS, TALCOTT 1951 *The Social System*. Glencoe, Ill.: Free Press.
- RADCLIFFE-BROWN, A. R. (1952) 1961 *Structure and Function in Primitive Society: Essays and Addresses*. London: Cohen & West; Glencoe, Ill.: Free Press.
- SCHNORE, LEO F. 1958 *Social Morphology and Human Ecology*. *American Journal of Sociology* 63:620-634.
- SELVIN, HANAN C.; and HAGSTROM, WARREN O. 1963 *The Empirical Classification of Formal Groups*. *American Sociological Review* 28:399-411.
- TANNENBAUM, ARNOLD S.; and BACHMAN, JERALD G. 1964 *Structural Versus Individual Effects*. *American Journal of Sociology* 69:585-595.
- VOGT, EVON Z. 1960 *On the Concept of Structure and Process in Cultural Anthropology*. *American Anthropologist New Series* 62:18-33.
- WEBER, MAX (1922) 1957 *The Theory of Social and Economic Organization*. Edited by Talcott Parsons. Glencoe, Ill.: Free Press. → First published as Part 1 of *Wirtschaft und Gesellschaft*.
- WILSON, GODFREY; and WILSON, MONICA H. (1945) 1965 *The Analysis of Social Change, Based on Observations in Central Africa*. Cambridge Univ. Press.

## SOCIAL SURVEYS

See EVALUATION RESEARCH; PANEL STUDIES; SAMPLE SURVEYS; SURVEY ANALYSIS.

## SOCIAL SYSTEMS

See under SYSTEMS ANALYSIS.

## SOCIAL WELFARE

See PHILANTHROPY; PLANNING, SOCIAL, *article on WELFARE PLANNING*; SOCIAL PROBLEMS; SOCIAL WORK; WELFARE STATE; and the articles listed under SOCIAL SECURITY.

## SOCIAL WORK

The objectives of social work are to help individuals, families, communities, and groups of persons who are socially disadvantaged and to contribute to the creation of conditions that will enhance social functioning and prevent breakdown. These objectives commit the social work profession both to helping persons adapt socially in keeping with their capacities and the norms and values of the society, and to modifying or reforming features of the social system. The term "social worker" refers to a special group among those employed in rendering social welfare services or conducting programs of agencies and institutions that make up the social welfare system. The professional social worker is expected, because of his specialized training and experience, to bring a high degree of skill to the process of helping, and modifying the social conditions of, individuals, groups of persons, and communities. The special competence of the professional social worker is exercised in such tasks as providing material assistance for the needy and dependent; assisting those of whatever means who have difficulties in adjusting to their economic and social environment because of poverty, illness, deprivation, conflict, or personal, family, or social disorganization; and participating in the formulation of social welfare policies and preventive programs.

### Emergence of the social work profession

As a deliberate concern for helping the indigent and less privileged members of society, the roots of social work go back to religious and humanitarian impulses evident in the histories of most civilizations. As a formal organization of efforts by specialized personnel to help such persons, social work is an accompaniment of nineteenth- and twentieth-century industrialization, with its associated problems of social dislocation from a more stabilized family and community system. The aspirations of positivist social science and the counterviewpoints to social Darwinism in the late nineteenth century contributed to the conviction that rational and scientific solutions to social problems could be found through social reform and individual guidance. The trend has been from personal and religious charity, to organized philanthropy, to public acceptance of



responsibility for programs of professional services. [See SOCIAL DARWINISM.]

As a profession, social work emerged during the twentieth century, when it became an acknowledged, full-time occupation, with established training schools, professional associations, and a high degree of self-consciousness about its status. Professionalization developed first in the United States and western Europe. Before 1910 independent schools for training social workers were founded first in the Netherlands and then in the United States, Great Britain, and Germany. Between World War I and World War II social work schools were widely established in Latin America and countries of the British Commonwealth. Since World War II, professional schools have opened in Asia, the Middle East, eastern Europe, and more recently in Africa. By 1950 a United Nations survey enumerated 373 schools of social work in 46 countries representing three-fifths of the 77 member nations. Since 1950 new schools have been founded in all regions, but especially in the new nations of Africa (United Nations 1950; 1965). Both national and international associations of social workers have been formed, professional journals and other publications issued, codes of ethical conduct promulgated, and official or legal recognition achieved to distinguish the professional social worker from other social welfare personnel.

Programs of social services under private and public auspices have expanded, and the number of social welfare workers has increased. The number of professionally trained social workers has also grown, but they still constitute a minority of all those engaged in social welfare work. In the United States in 1960, approximately one-fifth of the 105,000 social welfare workers had sufficient education to qualify as professional by the profession's own standards (*Salaries* . . . 1961). The proportion of professional to nonprofessional social welfare workers appears to be smaller in other countries. The profession thus faces a persistent problem of trying to establish and maintain an independent identity.

With the growth of educational facilities, the term "social worker" has increasingly come to refer to those specifically trained for social work, but the designation does not uniformly connote professional status in various countries. It sometimes refers to any sustained individual charitable efforts, and it usually includes employees, whether specifically trained or not, of public and private organizations serving the indigent or persons with problems of health, handicap, or deficiencies that may affect their social adequacy ("The Term . . .

1961). There is evidence that professionalization of social work is proceeding in similar form and with similar problems in various parts of the world. This is, in part, a result of the pressing demand for trained personnel and the diffusion of patterns from the more developed countries. Leaders in social work in developing nations often have had some of their training in western Europe, Canada, or the United States (*Interprofessional* . . . 1963), or have participated in international conferences in which professionalization has been prominently discussed.

Because most social workers are salaried employees who are characteristically dependent on bureaucratic organizations (generically called agencies), the development of a professional identity apart from the organization where they are employed has presented another persistent problem. The balance of influence over training schools has shifted gradually from agency to professional control. Moreover, there is a gradual tendency for professional associations to represent collective interests transcending those of social workers in particular agencies or fields of practice. Professional identification is also encouraged by frequent meetings and conferences of social workers on local, regional, national, and international levels.

### Demography of the social work profession

A number of characteristics of the personnel of social work bear on the nature of the profession. Women predominate, although the proportion of men has been increasing. In the United States in 1960, 60 per cent of all social welfare workers, but 68 per cent of the members of the professional association, were women. The proportion of female welfare workers in the United States decreased by nearly 10 per cent between 1950 and 1960, and the more balanced sex ratio of students in professional schools probably reflects the trend toward equality of the sexes in American society as a whole. In other countries, the preponderance of women is probably even greater, despite the efforts of professional and governmental bodies to recruit men into welfare work. Men tend to gravitate to supervisory and executive positions, but women nevertheless hold a considerable proportion of these jobs. This introduces sex-role conflicts within social work and between social work and other professions. As a "female profession," social work shares the generally lower prestige of women in the occupational world; and this, in turn, contributes to the difficulty of raising salaries in comparison with professions dominated by men. Career patterns and job mobility also tend to be differen-

dated by sex. Thus, a changing sex ratio may be expected to have a significant effect on the profession.

Although information on the social class of origin of social workers is limited, the profession appears to serve as a channel of upward mobility for both men and women (Pins 1963, chapter 3). Compared with other professions, social work has been relatively accessible to persons from minority groups and lower-class backgrounds. Social workers with professional schooling probably come from class positions that are higher than those of persons entering social work without special training but not as high as the level of persons entering professions such as medicine or law. Unlike these professions, social work clearly does not have the advantage of gaining prestige from the class background of its members. Rather, entrance into the profession often raises the class position of its members, and this probably affects both internal and interprofessional relations.

**Patterns of distribution and training.** The character of the profession is affected by the location of professionals within the social welfare system. Programs of income maintenance, such as support for dependent children and the aged and ancillary services to social insurance beneficiaries, employ the largest share of all welfare workers but the smallest proportion of those professionally trained. In the United States approximately one-third of all social welfare workers, but only 5 per cent of professionally educated social workers, were employed in public assistance programs in 1960 (*Salaries . . . 1961*). On the other hand, medical and psychiatric social work programs together employed 9 per cent of all social welfare workers but 29 per cent of those professionally trained. Professional social workers are also disproportionately to be found in noninstitutional child welfare programs. The consequence of such a differential distribution is to focus professional attention on a narrower range than that embraced by the social welfare system and to provide an unbalanced degree of professionalization among the various welfare programs. Only 3 per cent of public assistance workers in the United States in 1960 were professionally trained, whereas over half of the workers in medical and psychiatric programs were trained. This imbalance tends to promote viewpoints in the profession that define certain places of employment as more "professional" and others as less so. Such viewpoints underlie attitudes about suitable locations for professional practice. Similar tendencies have been noted in other countries.

The relative concentration of professionally

trained social workers in private, as distinguished from governmental, agencies has also affected the conception of attractive locations for professional employment. Approximately two-thirds of all social welfare workers in the United States in 1960 were employed by federal, state, or local governments, whereas only about half of the professional social workers were so employed. In other countries, where the private sector tends to be smaller—and usually church-supported—the attraction of private agencies does not appear to have been so great, although the absence of a large private sector is sometimes cited as a limitation on professionalization.

The disproportionate location of professional social workers in private agencies and in psychiatric, medical, and child welfare fields in the United States has historical explanations. The social workers most concerned about professional identity in the early decades of the century were those associated with fields in which clinical professions having greater prestige practiced, especially medicine and psychiatry. During the 1920s these social workers found psychoanalytic theories congenial as a basis for developing a distinctive mode of practice and supporting a claim to higher status among helping professions (Lubove 1965); the schools of social work largely accepted this viewpoint. When large-scale public assistance programs were developed as part of the social security system during the depression of the 1930s, they were primarily conceived as income maintenance programs to be administered by government employees rather than as programs requiring professional services. The charitably sponsored private agencies, now relieved of their traditional responsibility for economic aid, became especially receptive to the clinical orientation of trained workers who were also attracted to the clinical services associated with public programs in health, mental hygiene, and child welfare. The location of professional workers within the social welfare system has, in turn, affected recruitment and training, encouraging further emphasis on the clinical practices within social work. Nevertheless, professional social workers have often occupied supervisory and administrative positions in public welfare, and the profession has sought to influence public welfare policies through its associations and spokesmen.

The training and distribution of professional social workers in the United States are being further shaped by new demands on social work in the 1960s. The 1962 amendments to the Social Security Act explicitly emphasized social services to recipients as an objective of public welfare, giving



impetus to the upgrading of educational requirements for welfare workers. New programs involving broadly conceived community approaches to delinquency, poverty, mental health, and other social problems have encouraged recognition of functions for social workers other than direct services to clients; and the increasing integration of private agencies into such government-supported programs has reduced the distinction between private and public auspices. A greater proportion of students at social work schools has been attracted to group work and community organization. The literature on professional practice has reflected greater interest in the social sciences. Moreover, the problem of professional manpower has received increasing attention from the profession's leaders and government officials. In other countries, where the undersupply is even more acute and trained social workers have gravitated toward clinical locations, some comparable trends are evident. Especially in less developed countries, the practice of community development—proposing a comprehensive attempt to affect the complex of institutions of a total rural or urban community—has attracted professional social workers and has been emphasized in their training. [See COMMUNITY, *article on* COMMUNITY DEVELOPMENT.]

The shortage of trained social workers is the manpower condition that underlies the patterns of distribution in developed as well as underdeveloped countries (Conference . . . 1965). It has given rise to active recruitment programs in most countries; and it has stimulated attempts to distinguish levels among professional responsibilities requiring varying amounts of training, as in Great Britain (Great Britain . . . 1959). In the United States the shortage has generated interest in the education of subprofessionals and in various patterns for their use in social agencies along with more highly trained workers, thus requiring new conceptions of the practice roles of professional social workers. There has been wide recognition that social workers cannot be trained in large enough numbers under present conceptions of professional education to fill all social welfare positions, even when many positions are excluded as not requiring professional training. The manpower shortage also complicates the efforts of the profession to maintain or raise its entrance standards. Such standards are under pressure to broaden the definition of professional membership, acknowledge the alternatives to professional training for entrance to the profession, and adapt to such demands or face submergence within the large mass of non-

professional workers. The response of the profession to these pressures will determine the character of recruitment, training, and utilization of professional social workers (National . . . 1966).

### Professional training

Who is a professional social worker and how one enters the profession are variously, and sometimes ambiguously, defined by different countries. Personal commitment and activity in professional affairs may serve to distinguish "professional" from other social welfare employees in some countries. Elsewhere (France, for instance), a state license designates the professional social worker. In general, however, formal professional education is increasingly the primary channel of entrance.

**Schools of social work.** Schools of social work throughout the world may be roughly compared on the basis of their number and distribution, their sponsorship and location within the educational system, and the diversity of their financial support. Western Europe and North America have a great number of both private and public schools of social work; private schools typically receive subsidies from public funds through various governmental budgets. The location of social work training varies: in Great Britain it is usually part of the curriculum of social studies departments in universities and colleges; on the Continent there are separate technical institutes; and in the United States and Canada there are postgraduate divisions for social work education. In Latin America, Asia, the Middle East, and Africa schools of social work, though still few in number, have increased markedly since 1950; and private or church-sponsored schools are becoming integrated with those established within public educational systems. The Soviet Union is reported not to have specialized schools or an established profession of social work (Madison 1962), although correctional workers are specially trained educators.

The relationships between professional schools and their sources of support may affect the type and character of training. Standards are set by law or government in a few countries, such as France and Brazil. In the United States, the agencies of government that furnish subsidies for many students and some faculty salaries are usually interested in obtaining personnel for the established programs they administer and, hence, tend to favor training for direct, clinical services. In countries where there are associations of schools of social work, these generally seek a large voice in the determination of their own standards.

Schools may find it feasible to expand their capacities in those fields of practice already most professionalized, thus giving less emphasis to newer fields and broader social approaches. The problem of achieving balance has been recognized by both educators and governmental officials (Blackey 1964; Wittman 1965).

Completion of secondary schooling is generally the minimum requirement for admission to social work schools. In the United States and Canada, and in some schools of several other countries, the student must hold a baccalaureate degree in order to be admitted. Age limits, both lower and upper, are sometimes specified. Restriction to female students was at one time common, particularly in Latin American schools, but sex limitations are disappearing. Professed humanitarian values and commitment to service are personal qualities usually sought in applicants, and previous academic performance is always taken into account.

Throughout the world, most schools have relatively small enrollments, usually under one hundred. However, schools are expanding. In the United States in 1964 the largest school reported 409 full-time students, and 38 of the 59 schools reported full-time enrollments of one hundred or more, constituting approximately three-fifths of the 7,366 full-time students in all schools (*Statistics . . . 1964*, table 4). Social work schools also provide instruction, sometimes in separate courses, for many part-time students, only some of whom are enrolled for the professional degree; precise data are available only for the United States and Canada, where part-time students constituted approximately one-quarter of all students enrolled in the professional curriculum in 1964 (*Statistics . . . 1964*, table 2).

Students at most schools fall into two groups: those entering as an immediate continuation of their schooling and those entering after some period of employment in social welfare. In the United States the latter are a decreasing proportion, but increased public funds for training employed personnel may retard this decrease. Such students are generally older; thus their proportion in American schools may be estimated from the fact that in 1964 students 31 years of age or older constituted 30 per cent of all full-time students working for the professional degree (*Statistics . . . 1964*, table 9). Prior work experience is accorded decreasing importance relative to recency of academic study, but the accommodation of students with different backgrounds presents a characteristic problem in curriculum planning. The content of academic prereq-

uisites tends to be general—usually specifying only that courses in a number of social sciences be presented—and the academic backgrounds of students vary widely.

**Characteristics of training.** Social work education includes both theoretical and practical training. Practical training may be provided (as in Great Britain and countries it has influenced) by full-time field placements of students intermittent with classroom lectures or after completion of theoretical courses; or it may be provided concurrently with classroom work (as in the United States and countries influenced by it). Laboratory agencies for practical training under control of the professional schools are provided for a small proportion of students, but most are trained in selected, available agencies in the community. Practical field work generally requires "supervision"—that is, instructional direction by a trained social worker in an agency setting. A continuing difficulty in accommodating larger enrollments is the finding of suitable "field placements" for additional students. This system sometimes limits the range of agencies within which training may be given, particularly in new fields of practice, where few professionally trained social workers are employed.

The content of social work training tends to be organized around methods of social work practice, and field instruction is usually identified with one (but sometimes several) of the methods. The more theoretical portions of the curriculum provide knowledge of human behavior and social life deemed useful for competent practice. Courses on social welfare policies and the structure of welfare services of the country are usually required, and often some courses in social work research. Most students are trained in the casework method; in the United States in 1964 three-fourths of the full-time students were so identified. Casework method stresses interpersonal skills (usually the casework interview) to help individual clients or families solve problems of personal or social stress. The development of casework method has drawn heavily on psychoanalytic theory and dynamic psychology; and their perspectives pervade this training, although there is a marked trend toward the inclusion of a wider range of knowledge from the behavioral sciences. Training in a second method, called social group work, seeks to develop skills in using interaction processes of specially constituted or selected groups to help clients. A third method, called community organization, involves training to work with groups and organizations within the community in the solution of social problems.



Finally, a few students are given special training in social welfare administration or in social work research. Schools in developing countries have emphasized training in community development to a greater extent than have those in Europe and North America. However, community organization is currently being conceived more broadly in the United States, and it represents the most rapidly expanding part of the curriculum.

Along with more academic and technical training, social work schools usually give deliberate attention to the socialization of the student into the profession. The professional practitioner is presented as a model in most classes, but especially in field placements. Faculty advice and supervision by field work instructors tend to be close and fairly continuous through the years of study, and frequent evaluations of progress in acquiring professional norms, values, and orientations are often shared with the student. Strong professional identifications seem to lead students to prefer subsequent employment in positions where other professional social workers are practicing. Some observers of social work education in the United States have expressed concern with the academic level of graduate study and the focus of training on therapeutic rather than preventive content (Kadushin 1965). However, there is evidence that the social work curriculum is changing in terms of greater emphasis on social science materials and broader conceptions of the roles for which social workers are to be trained.

The number of years required for completion of social work training is related in part to the educational level attained by the students prior to enrollment in professional schools. If students are admitted directly from secondary schools, the length of study is usually three or four years. This generally corresponds in duration to the training given other technical professionals, such as teachers and nurses. The most significant exception to this widespread pattern is found in the United States and Canada, where the length of training is two years, but completion of four years of college is a prerequisite for admission.

Faculty members for social work schools are recruited primarily from practitioners in social agencies. Academic work beyond the professional degree has not generally been considered necessary. In recent years, as social work schools have become more integrated with universities, faculty members with advanced degrees have been sought; in the United States, professionals with doctorates in social work or in a social science are increasingly

employed. Fifteen schools of social work offered doctorates for advanced study of social work in 1964. Faculty members with graduate degrees in other academic disciplines—especially sociology and psychology—are often used, as are part-time instructors from other professional schools, especially from psychiatry, medicine, and law. While staffing patterns differ among countries, and a pattern of collaboration between practitioners and social scientists is fairly general, the point of view of the professional is usually dominant. There is widespread recognition of an acute shortage of personnel suitable for faculty membership.

### Professional practice

As previously noted, the profession of social work is almost exclusively practiced within agencies. Private practice, particularly of casework, has grown in the United States (Levenstein 1964), but it is not common. Social agencies range from those in which social workers constitute the exclusive or predominant professional personnel—such as family and children's service agencies, some community centers, and some social welfare planning councils—to institutions in which social workers occupy positions complementary or secondary to other professionals—such as hospitals, clinics, and public schools. Interaction between organizational and professional interests constitutes the continuing context for social work practice (Vinter 1959; Billingsley 1964).

The type of client and the kind of service offered are initially fixed by the purpose of the agency. A characteristic part of social work practice is "intake"—that is, determination of whether the client and the problem he presents can be properly accepted or should be referred to a different agency and, then, whether the client can be helped by the available professional treatment. In those instances in which clients are residents of correctional or medical institutions, pupils of schools, or cases before the courts, selection is further limited. The agency may also, by organizational decision as well as by the selective employment of social workers, determine the method of practice to be utilized. The introduction of different methods may be as much a decision of agency boards and executives as of professional practitioners. Although the executives are often social workers, practitioners are subject to evaluation in terms of agency criteria, such as productivity and conformity to routines, and these may not always be consonant with professional criteria of optimum service to clients.

There is always at least latent strain between

the requirements of professional autonomy and the constraints of the organization. Such strain is less evident where professional social workers occupy all levels in an agency. Where they are in a minority or subject to authority of other professions, the differentiation of a distinctive function for the professional social worker may become blurred, and its determination may become a matter of concern. Where untrained workers constitute a large proportion—as in most public welfare agencies—professional social workers often occupy supervisory or administrative positions. In agencies that provide mental health services—such as child guidance clinics, clinical and outpatient services of mental hospitals, and some community health agencies—social caseworkers (and sometimes social group workers) are often parts of clinical teams consisting of psychiatrist, clinical psychologist, and social worker. Problems of establishing appropriate roles and working out interprofessional relations have been observed, particularly those centering on the higher status often accorded to, or assumed by, the psychiatrist (Michigan . . . 1957).

The institution of "supervision" is characteristic of the organization of much social work practice. More experienced or more highly trained practitioners serve as "supervisors," not only in the administrative sense but as colleagues with whom practitioners are expected to discuss specific cases and problems of practice. This is viewed as a continuation of professional education for the social worker (usually on an individual, tutorial basis) and as a safeguard for standards of practice. Even trained social workers are supervised. This has been said to retard the achievement of professional independence, and it has recently come in for criticism and experimental modification (Burns 1965). No detailed empirical study of supervision in its organizational and professional context has yet been published, although it appears to be a common pattern for the organization of practice in most countries, it is often associated with the use of consultation about particular cases with other professionals, such as psychiatrists or other physicians, psychologists, and lawyers.

Social workers tend to identify themselves along lines of practice methods as caseworkers, group workers, or community organization workers, and their agencies are often similarly identified. There is experimentation, in both training and practice, with multiple uses of these methods and with their modification in order to find the most appropriate patterns of service. Particularly noteworthy are ef-

forts to devise new approaches to interpersonal helping processes, to the structuring of service and treatment organizations, and to the utilization of social action and other methods of community work. In the more clinical practices, such professional developments are strongest in the United States and western Europe, but new techniques in community organization are appearing most rapidly in other countries, where community development work in rural and urban areas is prominent.

**Professional-client relationships.** The major part of social work practice takes the form of face-to-face interviews with individuals, families, or groups of clients at the agency. Compared with other helping professions, however, social work conceives its practice as involving more contact with clients in their homes and elsewhere in the community. For example, work "in the streets" with gangs of delinquents is carried on by a number of agencies, and extensive work with neighborhood groups, settlement houses, and voluntary associations in the community is characteristic of some community organization practice. In addition, community organization practiced in welfare planning or fund-raising agencies calls for other relationships. The clientele of social workers thus consists of a wide range of persons, groups, and community organizations. Clients of public agencies are usually from lower socioeconomic levels as compared with the clients of private agencies (Cloward 1963). The latter may charge fees to clients who can afford them, but the services of social workers are usually provided without cost to the clients. Among the helping professions, social work also tends to adopt a more comprehensive conception of the social context of the client's problem. Hence, professional service often entails contacts with family members, employers, associates, and agencies that affect the client. Even so, critics both outside and within the profession frequently urge a wider social perspective, and such criticism is partly responsible for recent programs dealing with problems of delinquency, dependency, and illness in terms of the social environment.

The relationship between the professional social worker and his client is governed by strong norms of confidentiality and responsibility; it is seen as a privileged relationship delimited in function and content, calling for objective, emotionally neutral interest and impartiality in the rendering of service (Wilensky & Lebeaux 1958, pp. 298-303). Such norms are more explicitly recognized where professionalization is most advanced, but they are widely asserted in international literature as proper



for the professional social worker. Recognition of the professional nature of social workers by their clients is not always equally evident. This appears to encourage self-consciousness as well as concern for opinions of the staff and financial supporters who seem seriously concerned about their practice and image. The social worker views his practice as the use of his professional self to establish and maintain a helping relationship with his client, whether individual, family or community group. The centrality of the relationship is a basic value, as an indication that help is being given and thereby the worker's effectiveness is enhanced with the help they receive. However, in the United States less than half of the reports of private community agencies contain a second interview or contact with the client after discharge can be found in their case records, with the caring of being given over to the client.

Evaluation of social work. Importance of social work practice has long been recognized by the profession as a counterpart of its claim to professional competence. However, numerous evaluations of the effectiveness of social work have found problems ranging from those of policy, funding, and theoretical orientation. However, in all cases, many of the specific difficulties of the social work were never taken account of, except for the limited application to the evaluation of the social work itself, for example, in the evaluation of the effectiveness of the social work practice.

[illegible]

These findings are consistent with the hypothesis that the use of the word "and" in the sentence "John and Mary are married" is a marker of a specific type of relationship, namely, a relationship of equality. This is consistent with the findings of other studies which have shown that the use of the word "and" is a marker of a specific type of relationship, namely, a relationship of equality.

Marley & colleagues (1999) conducted a meta-analysis examining the impact of placement of children in foster care. All research had greater evidence for placement of children in foster care than in institutions or in orphanages. Marley & colleagues (2000) conducted a meta-analysis of 23 studies that examined the impact of placement of children in foster care. They found that placement of children in foster care had a positive impact on children's cognitive, emotional, and social development. They also found that placement of children in foster care had a positive impact on children's physical health. They concluded that placement of children in foster care is a positive intervention for children who are in need of a family.

strate only modest effectiveness of the theoretically oriented approaches utilized. Such findings encourage social workers to examine the theoretical bases for their methods and have contributed to a growing emphasis on manipulation of the social and community context in which the problems of clients arise.

Evaluation of the impact of community action programs presents complex problems that have only recently received systematic attention (see Freeman & Sherwood 1965). An attempt to affect delinquency rates by the use of street gang workers has been evaluated by Miller (1962), who found that such workers had little effect on delinquency rates. Others have evaluated the impact of community action (e.g. Ohlin 1960) or of community organization to enhance the position of the poor (e.g. Hagerstrom 1964); generate programs of such complexity that their rigorous evaluation is unlikely to be achieved until the modes of intervention are made more explicit and the results more clearly defined. The need for a more systematic approach to the evaluation of community action programs is thus apparent. The following review of the literature on the evaluation of community action programs is intended to provide a basis for the development of a more systematic approach to the evaluation of such programs.

career patterns. Another common factor in the career pattern of women is the fact that many women leave the profession after having interrupted their working life when they accompany their husbands to new jobs or when they become mothers. This is also one of the most characteristic differences between men and women social workers. There are also differences between men and women professional social workers that are partly a function of the disproportionate number of men in higher posts and in fields of practice, such as community organization work, which are higher (see Haker 1963 p. 537).

[illegible]

over involvement in public assistance agencies, causing disadvantaged parents, especially parents of inner-city youth.

### Professional associations

Many organizations to which social workers belong are termed the house of professional family or some similar phrase of interest, such as public and child welfare, law, medicine or mental health. Two types of strictly professional associations have emerged: the type that serves individual social workers as a membership association concerned with professional development in the profession, standards of professional practice, standards of educational preparation, ethics, and the status of social work toward the profession. The other type is the association of advocates of social work interests with policy and lobbying, educational requirements and standards, support for professional equity, and acceptance of formal status benefits. The former and members of these two types of associations assume, and then often share, the responsibility for professional equity, whereas the latter would regard it as their responsibility to represent themselves as the professional community or social workers as the community in practice, and affairs of discipline, equity, development and the status of social work. The distinction more recently, especially at schools, has been in that either be an *association* (United Nations, 1980). In Europe, where professional membership associations are rare and well-regulated, the association of schools play greater but still significant role in the United States with regard to discipline, equity, and status, but not membership. In Europe, where professional membership associations are rare and well-regulated, the association of schools play greater but still significant role in the United States with regard to discipline, equity, and status, but not membership.

Various international social welfare professional associations have formed since the turn of the century, but none is a majority of countries. However, organizations have proliferated, specializing first in the number of professions social workers, second in their geographic scope, and third, their social philosophies have culminated, enough to constitute but single national organizations. The National Association of Social Workers (NASW), which was organized in 1955, presently represents 100,000 members nationwide. For all social workers in the United States, with exception that recognized as one member, NASW remains a national-level association of social work professionals. In the United Kingdom, the Council on Social Work Education (CSWE) established in 1962, is

only for accreditation only those schools which grant the master's degree after two years of graduate study. It also maintains a division for undergraduate departments that give social work instruction regarded as "preprofessional."

The International Federation of Social Workers, founded in 1932 and reactivated in 1950 after the interruption of World War II, is composed of national associations of social workers. The International Association of Schools of Social Work has also been established. These organizations carry on informational and educational activities, meeting biennially along with the International Conference of Social Work, an organization with broad social welfare interests.

Membership in NASW accords recognition of professional status in the United States and Canada. A degree from an accredited graduate school is an eligibility requirement for certain civil service positions of state and federal governments. NASW has sponsored legal certification restricting the use of the title "social worker" to those having specialized certifications, and in 1980 four states had laws protecting the title. In France and Puerto Rico, social workers are licensed, and several countries have established legal requirements for schools of social work. In 1960, NASW founded the Academy of Certified Social Workers (ACSW) to designate social workers who are members of NASW and have two years of successful social work experience under the supervision of a member of the academy. The effect of ACSW will be, after a transition period, to establish a differential within the profession that is based on practice experience versus their education. Proposals to broaden the base of the profession by admitting social workers with college degrees and a certain number of years of experience to some form of membership in NASW were unsuccessful up to 1980. In the continuing debate over the issue, some members argue that present membership restrictions may generate competitive professional organizations with other qualifications for membership. It is more support the position of a "professional" rather than an "occupational" association. The issue may be expected to persist, especially in the face of the shortage of trained social workers and the influence of parents with qualifications other than social work training, who compete in social welfare under large public programs in such areas as mental health, poverty, delinquency, and community youth.

Associations of professional social workers in a number of countries have adopted or proposed



codes of ethics, and in the United States NASW promulgated such a code in 1960. It pledges members to a series of obligations, including primary responsibility to the client, confidentiality, recognition of the limits of professional knowledge and competence, appropriate respect for colleagues, priority of professional over personal interests, and other standards of professional conduct. There has also been discussion of an international code of ethics by the International Federation of Social Workers.

### Future directions

The character of the profession is being affected by several lines of development within social work. One of these is the increased emphasis on institutional and community change, in which the client is encouraged to participate. A number of governmental programs have adopted this approach, making the environment rather than the individual the major target of intervention. Social workers involved in these programs are turning to broader theories of social science in their training and practice and are enlarging the scope of community organization.

Social workers who help the client to function better in his individual situation are also experimenting with additional techniques. There is renewed consideration of the whole family, and treatment methods more commonly include the use of group approaches. Moreover, social workers are increasingly turning to behavioral psychology as a guide for practice rather than to the psychoanalytic doctrines upon which the therapeutic methods of social work have largely depended.

Manpower pressures are stimulating experimentation with new forms of practice and with the differentiation of tasks for which varied training is required. The increasing responsibility of social workers in the implementation of large public programs and greater security about their competence are encouraging both basic and evaluative research. Finally, achievement of more secure status and recognition is increasing the social work profession's ability to define the role it will play in the changing society.

HENRY J. MEYER

[Directly related are the entries PLANNING, SOCIAL, article on WELFARE PLANNING; WELFARE STATE. Other relevant material may be found in PHILANTHROPY; POVERTY.]

### BIBLIOGRAPHY

The absence of extensive empirical studies limits analysis of the social work profession. Useful descriptive materials on social work education and other aspects of the

profession can be found in the four United Nations surveys published under the title *Training for Social Work* in 1950, 1955, 1958, and 1965. Current developments in various countries are reported in *International Social Work*, published bimonthly by the International Conference of Social Work. See Great Britain . . . 1959 for an extensive study of social work in Great Britain, focusing on manpower and training; Timms 1964 provides a history of British social work in mental health; and Rodgers & Dixon 1960 presents an intensive study of social workers in a British community. For the United States, the *Encyclopedia of Social Work* contains a wide range of information; it was published in 1965 by the National Association of Social Workers, replacing an earlier series, the *Social Work Yearbook. Salaries and Working Conditions* . . . 1961 includes some data on trends in the United States since 1950. The emergence of the profession in the United States is analyzed historically in Lubove 1965. The most complete sociological analysis of the profession is a perceptive study of social welfare, Wilensky & Lebeaux 1958; other general analyses of the profession include Pollack 1952; Greenwood 1957; Cohen 1958; Meyer 1959. For studies of the characteristics of social work students, see Pins 1963 and Berengarten 1964. Manpower issues are reviewed in National . . . 1966. The Council on Social Work Education's publications include the *Journal of Education for Social Work*, regular informational bulletins, and an annual series entitled *Statistics on Social Work Education*. The National Association of Social Workers also has a publications program that includes two quarterly journals, *Social Work and Abstracts for Social Workers*, as well as regular bulletins on personnel information and professional developments. Professional journals, especially *Social Service Review*, *Social Casework*, and *Child Welfare*, frequently include discussions of problems of professionalization. Developments in public welfare bearing on the profession may be found in the journal *Public Welfare* and in other publications of the American Public Welfare Association, as well as in occasional publications of the Welfare Administration, U.S. Department of Health, Education and Welfare. In addition, more technical papers about social work knowledge and practice often supply insights into trends in the profession.

*Abstracts for Social Workers.* → Published since 1965.

BAKER, MARY R. 1965 Personnel in Social Work. Pages 532-540 in *Encyclopedia of Social Work*. New York: National Association of Social Workers.

BECK, DOROTHY F. 1962 *Patterns in Use of Family Agency Service*. New York: Family Service Association of America.

BERENGARTEN, SIDNEY 1964 *Admissions Prediction and Student Performance in Social Work Education*. New York: Council on Social Work Education.

BILLINGSLEY, ANDREW 1964 Bureaucratic and Professional Orientation Patterns in Social Casework. *Social Service Review* 38, no. 4:400-407.

BLACKKEY, EILEEN 1964 Issues in Social Work Education: New and Changing Demands Made of the Profession. Pages 75-89 in Council on Social Work Education, Twelfth Annual Program Meeting, *Proceedings*. New York: The Council.

BLENKNER, MARGARET 1950 Obstacles to Evaluative Research in Casework. *Social Casework* 31:54-60, 97-105.

BLENKNER, MARGARET 1962 Control Groups and the "Placebo Effect" in Evaluative Research. *Social Work* 7, no. 1:52-58.

BLENKNER, MARGARET; JAHN, JULIUS; and WASSER, EDNA 1964 *Serving the Aging: An Experiment in Social Work and Public Health Nursing*. New York: Community Service Society.

- BRIELAND, DONALD 1959 *An Experimental Study of the Selection of Adoptive Parents at Intake*. New York: Child Welfare League of America.
- BURNS, MARY E. 1965 Supervision in Social Work. Pages 785-791 in *Encyclopedia of Social Work*. New York: National Association of Social Workers.
- Child Welfare*. → Published since 1922.
- CLOWARD, RICHARD A. 1963 Social Class and Private Social Agencies. Pages 123-137 in Council on Social Work Education, Eleventh Annual Program Meeting, *Proceedings*. New York: The Council.
- CLOWARD, RICHARD A.; and OHLIN, LLOYD E. (1960) 1961 *Delinquency and Opportunity: A Theory of Delinquent Gangs*. New York: Free Press.
- COHEN, NATHAN E. 1958 *Social Work in the American Tradition: Field, Body of Knowledge, Process, Method and Point of View*. New York: Dryden.
- CONFERENCE ON INTERNATIONAL SOCIAL WELFARE MANPOWER, WASHINGTON, D.C., DEC. 13-14, 1964 1965 *Proceedings*. Washington: U.S. Department of Health, Education and Welfare, Welfare Administration.
- Encyclopedia of Social Work*. Edited by Harry L. Lurie. 1965 New York: National Association of Social Workers.
- FREEMAN, HOWARD E.; and SHERWOOD, CLARENCE C. 1965 Research in Large-scale Intervention Programs. *Journal of Social Issues* 21, no. 1:11-28.
- GREAT BRITAIN, WORKING PARTY ON SOCIAL WORKERS IN THE LOCAL AUTHORITY HEALTH AND WELFARE SERVICES 1959 *Report*. London: H.M. Stationery Office.
- GREENWOOD, ERNEST 1957 *Attributes of a Profession*. *Social Work* 2, no. 3:45-55.
- HAGSTROM, WARREN O. 1964 The Power of the Poor. Pages 205-223 in Frank Riessman et al. (editors), *The Mental Health of the Poor*. New York: Free Press.
- HUNT, JOSEPH McV.; and KOGAN, LEONARD S. 1950 *Measuring Results in Social Casework: A Manual on Judging Movement*. New York: Family Service Association of America.
- HYMAN, HERBERT H.; WRIGHT, CHARLES R.; and HOPKINS, TERENCE K. 1962 *Applications of Methods of Evaluation: Four Studies of the Encampment for Citizenship*. Berkeley: Univ. of California Press.
- International Social Work*. → Published quarterly by the International Conference of Social Work.
- INTERPROFESSIONAL CONFERENCE ON PROFESSIONAL TRAINING IN NORTH AMERICA FOR STUDENTS FROM OTHER COUNTRIES, OSGOOD HILL CONFERENCE CENTER, 1961 1963 *The Professional Education of Students From Other Lands*. Prepared by Irwin T. Sanders. New York: Council on Social Work Education.
- Journal of Education for Social Work*. → Published since 1965.
- KADUSHIN, ALFRED 1965 Two Problems of the Graduate Program: Level and Content. *Journal of Education for Social Work* 1, no. 1:33-46.
- LEVENSTEIN, SIDNEY 1964 *Private Practice in Social Casework: A Profession's Changing Pattern*. New York: Columbia Univ. Press.
- LUBOVE, ROY 1965 *The Professional Altruist: The Emergence of Social Work as a Career, 1880-1930*. Cambridge, Mass.: Harvard Univ. Press.
- MADISON, BERENICE 1962 Welfare Personnel in the Soviet Union. *Social Work* 7, no. 3:57-68.
- MEYER, HENRY J. 1959 Professionalization and Social Work. Pages 319-340 in Alfred J. Kahn (editor), *Issues in American Social Work*. New York: Columbia Univ. Press.
- MEYER, HENRY J.; and BORGATTA, EDGAR F. 1959 *An Experiment in Mental Patient Rehabilitation: Evaluating a Social Agency Program*. New York: Russell Sage Foundation.
- MEYER, HENRY J.; BORGATTA, EDGAR F.; and JONES, WYATT C. 1965 *Girls at Vocational High: An Experiment in Social Work Intervention*. New York: Russell Sage Foundation.
- MICHIGAN, UNIVERSITY OF, RESEARCH CENTER FOR GROUP DYNAMICS 1957 *Role Relations in the Mental Health Professions*, by Alvin Zander, Arthur R. Cohen, and Ezra Stotland. Ann Arbor: Univ. of Michigan, Institute of Social Research.
- MILLER, WALTER B. 1962 The Impact of a "Total-community" Delinquency Control Project. *Social Problems* 10:168-191.
- NATIONAL ASSOCIATION OF SOCIAL WORKERS, RESEARCH SECTION 1959 *Use of Judgments as Data in Social Work Research*. Edited by Ann W. Shyne. New York: The Association.
- NATIONAL ASSOCIATION OF SOCIAL WORKERS, RESEARCH SECTION 1966 *Manpower in Social Welfare: Research Perspectives*. Edited by Edward E. Schwartz. New York: The Association.
- PINS, ARNULF M. 1963 *Who Chooses Social Work, When and Why? An Exploratory Study of Factors Influencing Career Choices in Social Work*. New York: Council on Social Work Education.
- POLLACK, OTTO 1952 The Culture of Psychiatric Social Work. *Journal of Psychiatric Social Work* 21:160-165.
- Public Welfare*. → Published since 1943.
- RIPPLE, LILLIAN 1964 *Motivation, Capacity and Opportunity: Studies in Casework Theory and Practice*. Univ. of Chicago, School of Social Service Administration.
- RODGERS, BARBARA N.; and DIXON, JULIA 1960 *Portrait of Social Work: A Study of Social Services in a Northern Town*. Oxford Univ. Press.
- Salaries and Working Conditions of Social Welfare Manpower in 1960*. 1961 New York: National Social Welfare Assembly.
- SHYNE, ANN W. 1963 Evaluation of Results in Social Work. *Social Work* 8, no. 4:26-33.
- SHYNE, ANN W. 1965 An Experimental Study of Casework Methods. *Social Casework* 46:535-541.
- Social Casework*. → Published since 1920.
- Social Service Review*. → Published since 1927.
- Social Work*. → Published since 1956.
- Social Work Yearbook*. → Published from 1929 to 1960.
- Statistics on Social Work Education*. → Published since 1952.
- The Term "Social Work" as Used Throughout the World. 1961 *International Social Work* 4, no. 1:5-9; no. 3:29-32; no. 4:9-11.
- THOMAS, EDWIN J.; and McLEOD, DONNA L. 1960 *In-service Training and Reduced Workloads: Experiments in a State Department of Welfare*. New York: Russell Sage Foundation.
- TIMMS, NOEL 1964 *Psychiatric Social Work in Great Britain: 1939-1962*. London: Routledge; New York: Humanities.
- UNITED NATIONS Training for Social Work. → Four international surveys published in 1950, 1955, 1958, and 1965.
- VINTER, ROBERT D. 1959 The Social Structure of Service. Pages 242-269 in Alfred J. Kahn (editor), *Issues*



in *American Social Work*. New York: Columbia Univ. Press.

VINTER, ROBERT D. 1963 Analysis of Treatment Organizations. *Social Work* 8, no. 3:3-15.

VINTER, ROBERT D.; and SARRI, ROSEMARY C. 1965 Malperformance in the Public School: A Group Work Approach. *Social Work* 10, no. 1:3-13.

WILENSKY, HAROLD L.; and LEBEAUX, CHARLES N. 1958 *Industrial Society and Social Welfare: The Impact of Industrialization on the Supply and Organization of Social Welfare Services in the United States*. New York: Russell Sage Foundation. → A paperback edition was published in 1965 by the Free Press.

WILKINS, LESLIE T. (1964) 1965 *Social Deviance: Social Policy, Action, and Research*. Englewood Cliffs, N.J.: Prentice-Hall.

WITTMAN, MILTON 1965 Information on Personnel Needs and Social Work Students: Implications for Manpower Planning and Research and for Programs of Recruitment and Education. Pages 15-25 in Ellen Winston et al., *Social Work Education and Social Welfare Manpower: Present Realities and Future Imperatives*. New York: Council on Social Work Education.

## SOCIALISM

The word begins in obscurity. Though various origins have been suggested, the first use in French has commonly been ascribed to the *Globe* of February 13, 1832, where the word *socialistes* was chosen to describe the followers of Saint-Simon. (However, a recondite reference to socialism a year earlier, in the religious journal *Le semeur*, has been uncovered.) Englishmen have claimed the honor of its coinage, since the word "socialist" did appear in the *London Co-operative Magazine* in 1826, although it was not until several years later that followers of Robert Owen began describing themselves as socialists. Clearly, however, the term was in the air, for it described a converging mood; and the first article on "socialism" as an idea in opposition to "individualism" was written by Pierre Leroux and appeared in 1835 in the *Encyclopédie nouvelle*, edited by Leroux and Reynaud. The word recurred thereafter in various writings by Leroux.

By 1840 the term "socialism" was commonly used throughout Europe to connote the doctrine that the ownership and control of the means of production—capital, land, or property—should be held by the community as a whole and administered in the interests of all. Within 120 years after the term became known in Europe, the doctrine had spread so widely that one could find regimes in Sweden, Great Britain, the Soviet Union, China, eastern Europe, Cuba, Algeria, Egypt, Syria, Israel, Guinea, Kenya, Tanzania, India, Burma, and Ceylon calling themselves socialist, and the labels

Arab socialism, African socialism, and Asian socialism used to describe the grafting of indigenous traditions onto ideological doctrine. Rarely in the history of the world has an idea taken hold so deeply and dispersed so quickly. One would have to go back to the spread of Islam, in the century and a half following the death of Muhammad, to find a comparable phenomenon. And the analogy is not without relevance, for one finds in both instances the promise of a perfect community, the effort to create a solidarity larger than that of tribe or class, a reaction to the meaninglessness of existing religious beliefs, a militant proselytizing spirit, and leadership by new elites. In fact, the comparison with Islam is meant to suggest that the spread of socialism cannot be wholly accounted for in economic or class terms. The socialist movement has (or had) the character of a secular religion, and only from this view can one explain its development and internal vicissitudes.

This article will discuss the formulation of early socialist doctrine, the differentiation of the socialist movement and the spread of socialism, the role of socialist parties, and varieties of socialist belief since Marx. [*Marxist views and positions are elaborated in MARXISM and in the articles under COMMUNISM.*]

### Growth of the doctrine

**Beginnings.** The meaning of socialism, both logically and sociologically, can only be understood as a contrast to individualism. The Enlightenment, English political economy, the French Revolution, and the nascent industrialism had all combined to produce what in 1826 a disciple of Saint-Simon called individualism. In this doctrine, society existed to serve the individual and the pursuit of his own satisfactions; natural rights inhered in each individual, and government was not to regulate the economic life of society. Even the French Revolution, with all its passion for virtue and its defense of popular sovereignty, fostered the idea of economic individualism.

The attack on individualism drew its strength from a Catholic and a socialist point of view. Bonald and de Maistre, both theocrats, were militantly against "political Protestantism" and asserted that man exists only for society. Particularly after the revolution of 1830, many French writers of a conservative bent—Lamartine, Balzac, Sainte-Beuve, Lammenais, and Tocqueville—expressed their alarm about *l'odieux individualisme* and held it responsible for the disintegration that they felt was occurring in their society.

While the conservatives attacked the political

philosophy which they linked to the French Revolution, the socialists were appalled by the economic doctrine of laissez-faire: this, Louis Blanc declared, was responsible for man's ruthless exploitation of man in modern industry. Under industrialization, the socialists alleged, the individual had been torn from old moorings and had no anchorage. Friedrich Engels, writing about London in *The Condition of the Working Class in England*, described "the brutal indifference, the unfeeling isolation of each in his private interest," which people experienced in the British capital, and stated that "the dissolution of mankind into monads of which each one has a separate principle, the world of atoms, is here carried out to its utmost extreme" ([1845] 1958, p. 24).

Against the atomization and "egoism" of society, as Saint-Simon called it, the social critics proposed a new order based on association, harmony, altruism, and, finally, the word that superseded all of these—socialism. The idea of socialism has a long history in the utopian tradition; one can trace its roots back to the dream of returning to a golden age of social harmony or to the radical theological creed—expressed most vividly by the Anabaptists of the sixteenth century and the Levellers and Diggers of the seventeenth—of the equality of all men. But equality alone is not the essence of socialism. The heart of socialism is to be found in the idea of community and in the doctrine that men can realize their full potential and achieve human emancipation in community. By this touchstone, the seeds of modern socialism are to be found in Rousseau [see the biography of ROUSSEAU].

The theme of community is also the central theme of Fourier, Owen, Saint-Simon, and Marx. The first three sought to achieve it through the a priori elaboration of the theoretical elements of community. Marx, on the other hand, sought to realize it through the sphere of philosophy and what he held to be its material embodiment, the proletariat. It is in the phrase "the realization of philosophy," the end point of a process of history, and not in any alleged distinction between utopian and scientific descriptions of socialism, that the difference between Marx and the others lies.

Both Owen and Fourier sketched socialist utopias that were enormously attractive to individuals whose sensibility was repelled by the evils of industrialism. Each wanted to establish a small agrarian community that science could make practical—in effect, a withdrawal from society. Neither man had a sense of history or any realistic awareness of the politics of his time. [See the biographies of FOURIER and OWEN.]

Saint-Simon ("the last gentleman and the first socialist" of France) was a very different sort, and the customary inclusion of him with Fourier and Owen as a "utopian" is actually a disservice to a formidable intellectual, a disservice initially performed by Marx, who, although he derived many ideas from Saint-Simon, failed to see the implications of much of the French writer's thought. John Stuart Mill, however, clearly recognized Saint-Simon's contribution, remarking, in *Principles of Political Economy* (1848), that in the few years of its public promulgation, Saint-Simon's thought had sowed the seeds of nearly all the socialist tendencies. Durkheim considered Saint-Simon to be the father of socialism, as well as of positivism, and devoted a book to his theories. Although in the *Communist Manifesto* Marx cavalierly dismissed Saint-Simon as a utopian, Engels in his later years remarked that Saint-Simon's "breadth of view" and "genius" contained in embryo "all the ideas of later socialists which are not strictly economic." For what Saint-Simon presented is what we know today as the theory of industrial society, and his discussion of the nature of solidarity outlines the theory of occupational community which Durkheim later elaborated. [See the biography of SAINT-SIMON.]

It is not too much to say, following Markham (1952), that the Saint-Simonians were the most important single force behind the great economic expansion of the Second Empire, particularly in the development of banks and railways. Enfantin, the most bizarre of the Saint-Simonians, formed the society for planning the Suez Canal. The brothers Émile and Isaac Pereire, who promoted the first French railway from Paris to Saint-Germain, also founded the Crédit Mobilier, the first industrial investment bank in France, and the Compagnie Générale Transatlantique, whose first ships were named after Saint-Simon and his followers.

In the hands of some of his more zealous followers, Saint-Simon's doctrines were made to seem ludicrous. Yet his own insight was considerable, and it was the Saint-Simonians' more diffuse (but no less intense) belief in Marxism which gave that doctrine its command over so large a part of the world.

**Marxism.** The *Communist Manifesto* (Marx & Engels 1848) and the writing done in the thirty years following it make up the corpus of work that later socialists drew upon and associated with Marx [for a detailed discussion, see the biographies of ENGELS and MARX]. Relying on the political activities of Marx as well as on his judgments, the



diverse socialist factions sought to justify their own policies. Thus Lenin and the Bolsheviks found in the address to the Communist League of March 1850, and in Marx's *Critique of the Gotha Programme* (Marx & Engels 1875-1891), the justification of their revolutionary and insurrectionary tactics. From Marx's activity in Cologne in the early part of 1849 and from his inaugural speech to the Grand Council of the International Workingmen's Association (the First International), democratic socialists have argued that peaceful electoral change is possible in the achievement of socialism.

Marx envisaged a two-stage development in industrial countries that presaged the victory of socialism. The first was the democratic revolution; the second, the social revolution. By the "democratic revolution" Marx meant the victory of the middle classes over the remnants of the aristocracy and the clearing away of feudal remains to achieve the successful development of capitalist production and of political rights for all in the society. By the "social revolution" Marx meant the economic victory of the proletariat, who will take over the ownership of the means of production. From this point of view, England was the most advanced country and, therefore, presumably the one most ripe for socialism. Measured on the same yardstick, Germany, while developing industrially, was still lagging behind because the middle classes had failed to complete the democratic revolution; Russia was the least advanced, since the middle classes had failed to act at all, whereas the German middle classes had at least done something in 1848. In the socialist perspective, therefore, Russia was the country where a revolution was most imminent, since it had lagged so far behind and was only beginning to catch up economically. But before 1914 all orthodox Marxists expected this to be a bourgeois, rather than a socialist, revolution, since the working class in Russia was too small to sustain a socialist revolution and the economy was too "immature" for socialism. Against the populists and the anarchists in Russia who argued that the country could skip a social stage and usher in a socialism based in part on the old village *miry*, or communal holdings, the early Russian Marxists—Axelrod, Plekhanov, and Lenin—argued that socialism in Russia would have to await the development of capitalism and the creation of a sizable working class. Only at the beginning of the twentieth century did the thought occur to some Russian Marxists, notably "Parvus" and Trotsky—and later Lenin himself, when he was converted to the idea—that they could use the impending Russian

revolution to wrest power from the bourgeoisie and thus spark revolution in the advanced industrial countries. Before 1917, no Marxists thought that socialism would be possible in preindustrial or underdeveloped countries. The West was expected to lead the way.

What kind of socialism was supposed to emerge? What would society be like the day after the revolution? And what were Marxists supposed to do while waiting for the revolution? No political party can exist without a program that holds out the promise of immediate benefits. But as Schumpeter has pointed out, anything positive done or to be done in the vitiated atmosphere of capitalism was *ipso facto* tainted. Marx and Engels discouraged programs that involved constructive policy within the capitalist order because they smacked of bourgeois radicalism. However, when they faced the problem in 1847, they resolutely cut the Gordian knot. As Schumpeter put it: "The *Communist Manifesto* quite illogically lists a number of immediate objects of socialist policy, simply laying the socialist barge alongside the liberal liner" (1942, p. 317 in the 1962 edition).

The problem was to recur constantly throughout the political history of most of the European socialist parties. Should one make immediate demands or not? This issue was fought out, for example, within the American Socialist party at the turn of the century; and it resulted in such factions as the Reformists, and the Impossibilists, who declared themselves against any such program on the ground that it would dilute the revolutionary ardor of the masses. More important, the problem of reforms, and of what kind of reforms, had to be confronted by the various socialist parties of Europe in the 1930s—such as the British Labour party, the German Social Democratic party, and the French Socialist party—when they entered the government and even took over sole responsibility for running it in a capitalist society. As we shall see, many of these governments and countries foundered when the socialist governments discovered, for example, that *they* had no solution for the problem of unemployment under capitalism. The slogan "Socialism or Capitalism" had left them unprepared for the exigencies of the intermediate period. This is always the dilemma of social movements that live in a world but are not of it.

Gotha and Erfurt. Marx did, of course, distinguish between socialism and communism, in the sense that the first is a transitional period and the second the undefined realm of man's freedom. For this aspect of the idea of socialism, two documents are crucial—the *Gotha Programme* (see Marx &

Engels 1875-1891) and the *Erfurt Program* (see Kautsky 1892), two doctrinal statements of the German Social Democratic party.

In 1875, two German socialist parties, one dominated by followers of Ferdinand Lassalle, the other by Wilhelm Liebknecht and August Bebel, nominal followers of Marx, met in Gotha to create a unified socialist party. The program they adopted there was in the main a product of Lassallean doctrine, and in a private communication Marx wrote a searing critique of it. In 1891, the German Social Democratic party adopted the so-called *Erfurt Program* (named for the city where the party congress was held), which was written principally by Karl Kautsky, under the direct supervision of Engels; Engels then published the *Critique* for its historic interest, feeling that the *Erfurt Program* went beyond the criticism Marx had made of the Gotha document. But the *Critique*, in its tone and implications, was more revolutionary and radical than the *Erfurt Program*, and, predictably, the left-wingers in the socialist movement, beginning with Lenin, formulated their program from the *Critique*. It was in the *Critique* that Marx used the phrase "revolutionary dictatorship of the proletariat," which after 1917 provoked far more recriminatory debates between communists and socialists than any other words he wrote. The communists took it as a justification of the suppression of all political parties other than the true proletarian party under Soviet rule. The socialists insisted that the phrase applied only to temporary and special situations until a truly democratic state could be organized, and was nugatory in those situations where a peaceful transition to socialism, through the electoral process, was possible.

The inflammatory phrase does not appear in the *Erfurt Program*, nor does any statement of immediate aims or political demands. The program was intended as a full-fledged analysis of the tendencies of capitalism, from a Marxian point of view, and as a general discussion of the "cooperative commonwealth" of the future. The program assumes the recurrent Marxist theme: "Few things are . . . more childish than to demand of the socialist that he draw a picture of the commonwealth which he strives for. . . . Never yet in the history of mankind has it happened that a revolutionary party was able to foresee, let alone determine, the forms of the new social order which it strove to usher in" (Kautsky [1892] 1910, pp. 122-123).

On the forms of organization, the managerial problems of a socialist regime—how orders are to be given, who will give them, which industries are to be managed by workers directly, which by state

enterprises; in short, the practical problems that the Soviet state faced after the communists had assumed power—the program is completely silent.

Kautsky, who inherited Engels' mantle as the leading Marxist theoretician, was prompted only once to deal with the problem of the organization of production in a socialist society (but not with the structure of authority within an enterprise). In some lectures delivered and published in 1902 as *The Social Revolution*, he declared simply that the organization of production would follow the scope of the market.

For example, gas lighting is clearly a municipal business. The development of electric lighting and the transformation of power in mountainous regions makes the nationalization of water power necessary. This operates also to transform illumination from a municipal to a national business. Again the business of the shoemaker was formerly confined to the local market. The shoe factory does not supply simply the community, but the whole nation, with its production, and is ripe not for communalization, but for nationalization. The same is true of sugar factories, breweries, etc. (1902, pp. 115-116)

In fact, when Kautsky had finished with his itemization (transportation, railroads, steamships, mines, forests, iron foundries, machine manufactures), it was clear that almost all industries would be nationalized in the "proletarian regime." [See the biography of KAUTSKY.]

If one goes beyond these pedestrian problems, however, it is interesting that the *Erfurt Program* ends, curiously enough, on a note reminiscent of the young Marx and of that strain in German romanticism which looked back to the glory of Greece.

The blessed harmonious culture, which appeared only once in the history of mankind and was then the privilege of a small body of select aristocrats, will become the common property of all civilized nations. What slaves were to the ancient Athenians, machinery will be to modern man. Man will feel all the elevating influences that flow from freedom from productive toil, without being poisoned by the evil influences which, through chattel slavery, finally undermined the Athenian aristocracy. And as the modern means of science and art are vastly superior to those of two thousand years ago, and the civilization of today overshadows that of the little land of Greece, so will the socialist commonwealth in moral greatness and material well-being the most glorious society that history has thus far known. (Kautsky [1892] 1910, p. 158)

#### Differentiation of the movement

The period from 1870 onward in western Europe saw the swift growth of industrialization and urbanization, the two crucial elements of modern



society. This expansion of industrial power and of economic growth and wealth, which was due largely to two technological innovations—the improvement of steel metallurgy and the application of electrical energy to factory, city, and home—seemed to confirm a number of Marx's predictions regarding the development of capitalism (see Marx 1867–1879, vol. 1, chapter 25, and vol. 3, chapter 23). Capitalism was undergoing remarkable changes. The expansion of the joint stock company (the prototype of the modern corporation) was forcing a separation of ownership and management, which in many areas resulted in the industrial manager's taking the place of the capitalist as the central person of the organization, and the large-scale enterprise began employing hundreds and even thousands of workers under a single roof. More important, the “amalgamation” movements of the 1880s and 1890s—the rise of trusts, cartels, and monopolies—and the consequent elimination of hundreds of smaller businesses seemed to bear out Marx's predictions about the centralization of capital and the socializing of the processes of production.

Volume 1 of *Capital* was published in 1867, and the subsequent expansion of the volume, along with its rapid translation into many languages, gave Marx, hitherto a neglected and cantankerous émigré in London, an authority in the international socialist movement, particularly in its German branch, which he had never had before. With the assiduous publication and spread of Marx's works by the growing socialist movements, Marxism suddenly became a vogue as no other socialist doctrine had ever been; and with the proliferation of followers and propagandists who in newspapers, pamphlets, and street meetings proselytized the simplified works of Marx, the doctrine itself assumed a canonical status that was unprecedented in the history of secular writing.

**The Second International.** In 1889 almost four hundred delegates from twenty different countries (three-fourths of them from Germany and France) met in Paris to create a new International, the Second International of socialist parties. The so-called First International, the International Workingmen's Association, was a loose confederation of small political and trade union groups, rather than parties, that had been organized in 1864. Although Marx was not the initiator of the First International, he quickly became its dominant intellectual figure, supplanting Mazzini, who had been asked to write its first draft program. The International broke up in 1872, when Marx and the anarchist leader Bakunin quarreled; though the anarchists

were expelled, Marx had the International's center moved to New York, preferring to bury it rather than allow some other group to capture it. The First International was formally dissolved in Philadelphia in 1876.

More than any other step, the founding of the Second International symbolized the swift rise of Marxist socialism in Europe. It was only 14 years earlier, in 1875, that the German Social Democratic party, the first socialist party in Europe, had been formed. In the next dozen years or so, socialist parties were organized in France, Austria, Belgium, Switzerland, Denmark, and Sweden. In Russia in 1883, the year of Marx's death, Georgii Plekhanov organized the first political group of Russian Marxists. About the same time, in England, M. H. Hyndman, the son of an aristocrat, organized the Social Democratic Federation, which, while calling itself Marxist, never acquired more than a small sectarian following; and a quixotic band of reformers organized the Fabian Society (the name alluded to the Roman general Fabius Cunctator—Fabius the Delayer—who was known for his patient, waiting tactics against Hannibal). In 1889, the year the Second International was founded, the historic *Fabian Essays in Socialism* was issued, with chapters by George Bernard Shaw, Sidney Webb, and Annie Besant. The book eventually sold two million copies, and laid the intellectual foundations of the British Labour party and of Labour governments for the next sixty years.

**Socialist parties.** By 1914, socialism had become the single most important political force on the Continent. In the 1912 Reichstag elections, the German Social Democrats amassed 4.5 million votes (over 30 per cent of the total) and 110 seats in the parliament, making it the largest single party in Germany. In France one of the socialist groups, the SFIO, garnered 1.4 million votes and 103 seats in the Chamber of Deputies. In Italy the socialists held over seventy seats in the parliament, and efforts were made to invite the party, or at least its right wing, into the government.

But the rise of the socialist parties was not only a simple matter of winning large numbers of votes, primarily among the working class. Within a new and growing system of universal political suffrage, it transformed the nature of the party system and the political structure of each country.

The political party of the first half of the nineteenth century was usually a loose association of “notables,” in Max Weber's terminology, invariably based on individual constituencies or districts, and often with little responsibility to an electorate.

With the growing democratization of the franchise in England, associations were formed in each district; and the caucus system, developed in 1868, enabled the Liberals to begin building local machines with full-time election workers. Yet mass membership was infrequent, and the parties of England, as well as the United States, depended for their finances on wealthy contributors. What the socialists did, particularly in Germany, was to introduce the disciplined and centralized mass party, with formal machinery for enrollment, regular payment of dues, a system of subscription to party newspapers and magazines, and, often, specified requirements of party activity. At its prewar peak, the German Social Democratic party had a million members and an annual budget of nearly two million marks.

But the socialist movement did more than build the first mass political party. It tried, in most of the European countries and to a lesser extent in England, to build a complete working-class culture, a social world of its own, independent of the official culture of the society. The German socialist movement, the model for all other socialist parties, built large consumer cooperatives (with a large wholesale organization and its own processing plants) as well as housing developments. By the 1890s there were national organizations of workers' athletic societies, workers' bicycling clubs, and workers' hiking clubs. In time, the workers' recreational and cultural movement extended into all fields from chess to the theater, where a strong *Volksbühne* (people's stage) was created. A working-class child could begin life in a socialist crèche, join a socialist youth movement, go to a socialist summer camp, hike with the socialist *Wandervogel*, sing in a workers' chorus, and be buried by a socialist burial society in a socialist cemetery.

If the idea of proletarian mountain climbing or socialist chess playing invites ridicule, one must see, as Carl Landauer (1959) points out, that the workers had been "excluded" from almost all accepted society, and they responded by creating their own.

**Revisionism and reformism.** The socialist movements at the turn of the century may have felt sure about inheriting the future, but there was considerable uncertainty as to when and how that inheritance would be realized. Marx, in all his writings, had never been specific about the road to power. After 1850 he felt that the day of the barricades was finished, not only for military reasons but also because bourgeois society would stabilize itself for a long time to come. Against this view, apocalyptic hopes occasionally flared up, as during

the Paris Commune. Yet Marx never took a dogmatic view as to any single course which the socialist movement would necessarily have to follow. In several instances, he felt that socialism might be achieved peacefully in the Western countries, where democratic institutions were being established. But he never ruled out the possibility of, and even the need for, violence, should the occasion demand it. Marx and Engels, throughout their lifetimes, insisted simply on the necessity of a revolution, by which they, as well as Kautsky, who became the leading spokesman for orthodox Marxism after the 1890s, meant a complete overturn of society once the socialists were in power—the abolition of private property, the end of social privilege, the breaking of the political and police power of the old ruling classes.

But the question whether this aim could be achieved by peaceful means was never settled. And this ambiguity was responsible for the major doctrinal conflicts that preoccupied the socialist movements from 1890 to 1914.

The major issues had to do with the themes of revisionism and reformism. Although their belief in socialism was never shaken, some individuals were skeptical that capitalist society was actually heading in the direction Marx had predicted. The standard of living was evidently rising rather than falling, and though some of the old middle class was disappearing, an emerging class of white-collar workers was taking its place. In many countries this new class did not wholly identify itself with the manual workers (with whom socialism was identified) or with the socialist parties. Most of all, the socialists' increasing success in parliament posed practical problems, such as entering the cabinets in coalition with other parties (and trying to put through social legislation rather than just waiting for capitalism to fall) and making alliances with nonworking-class parties such as the Liberals in England, the Catholic Center in Germany, or the Radical party in France. As James Joil has neatly put it: "By the end of the nineteenth century, no Socialist party could escape the difficulties presented by its own existence as a mass party, forced, for the moment at least, to function within a political system which at the same time it was seeking to destroy" (1955, p. 77).

**Germany.** The problem was especially great in Germany, whose Social Democratic party was the most theoretically intransigent, and it was first posed by Eduard Bernstein, who was the editor of the party journal and was chosen by Engels to be one of his literary executors. In 1899, Bernstein wrote *Die Voraussetzungen des Sozialismus und*



*die Aufgaben der Sozialdemokratie* (The Presuppositions of Socialism and the Tasks of Social Democracy), which triggered the debate. He argued, in effect, that the party should recognize the new changes in industrial society and declare itself to be what it was actually becoming—a party principally concerned with social reform. [See the biography of BERNSTEIN.]

Although Bernstein's arguments had some immediate political implications, the debate was conducted largely on a theoretical level. There were important doctrinal and historical reasons for this. The German movement had always set great store in theory as the guide to the future; and the discussion of theory—especially in a party isolated from the academic world—was in Germany a matter of status and prestige. But the debate on theory had important psychological roots as well. Although Marx had always tried to shape the course of the German socialist movement, his influence during most of his life was virtually nil. The first German Workers party was organized in 1863 by Lassalle, whose dandified manners and aristocratic pretensions infuriated Marx. The constitutional struggle between Bismarck and the liberal opposition dominated the political life of Prussia at the time; and to Marx and Engels, observing the situation from England, this struggle was an outward expression of the conflict between the historic forces of aristocratic feudalism and bourgeois capitalism. They therefore urged the new German Workers party "to drive the 'revolutionary' Liberals forward against the government, preparing at the same time to lead the proletariat in its turn against the victorious forces of the bourgeoisie once the feudal system had succumbed to their onslaught" (Morgan 1965, p. 8). But Lassalle and his followers were convinced that Prussian liberalism had no such revolutionary propensities and that the quickest way for the workers to increase their influence was to join Bismarck's campaign against his liberal opponents, hoping to win in return certain socialist demands—freedom of the press and association and, above all, manhood suffrage, which Lassalle and his followers felt was necessary to any further advance. Lassalle, under the influence of Louis Blanc, also hoped that the state would finance co-operative factories so that the workers could become their own employees and overcome the "iron law of wages" which kept them impoverished under capitalism. Lassalle, a Hegelian, believed that the state should rule society; and socialism, with state ownership of factories, was the embodiment of the ideal of the state. Against Lassalle (and his successor J. B. Schweitzer), the socialists

Liebknicht and Bebel organized a rival socialist party, known as the Eisenachers. Nominally Marxist, the Eisenachers were primarily an anti-Prussian party, and though they adopted the program of the International Workingmen's Association, they did so largely for tactical reasons against the Lassalleans. In practice they operated as a broad "people's party." The war of 1870 and the subsequent unification of Germany dissolved most of the issues between the two factions.

The debate remained on a theoretical plane for good practical reasons. Reformism—which assumed that effective parliamentary power could be obtained—was impossible in Germany; for while imperial Germany had a parliamentary system, decisive power was actually in the hands of the emperor. Any chancellor needed the Reichstag's approval for legislation and for the budget, but a chancellor could be replaced only if he lost the confidence of the emperor. The chancellor, in fact, was responsible not to the Reichstag but to the crown. The government, though constitutional in form, was in fact autocratic. And while social democracy was growing in parliamentary strength, there was no corresponding growth in the powers of parliament.

Thus the political system barred the Social Democrats from any legitimate hopes of winning power through parliamentary means and reinforced the rhetoric of revolutionary intransigence. The "fatal mistake," as Schumpeter called it, was Bismarck's. In a Machiavellian stratagem, he introduced universal suffrage in the federal empire after 1871, in the hope of winning the peasant votes, and to some extent those of the workers, against the urban middle classes. When the socialist vote began to increase, Bismarck introduced restrictive legislation against the socialists as a party, while introducing a comprehensive set of social welfare measures in order to win the loyalty of the workers. The maneuver failed. When the antisocialist laws lapsed in 1890, the socialists emerged stronger than ever, and their experiences reinforced their antagonistic temper and revolutionary rhetoric.

By the turn of the century, and thereafter, the party controlled large municipal administrations, was supported by a powerful trade union movement, and had a vast bureaucracy of its own. The party's ideology no longer corresponded to its sociological reality (Michels 1911), but, because the political system did not allow the party to discard its revolutionary rhetoric, the ideology remained intact.

Thus, when Kautsky came to answer Bernstein, he couched his polemic in the language of Marxist

scholasticism. In practical fact, the situation confronting the Social Democratic party of Germany—as well as the socialist parties of the Austro-Hungarian Empire and czarist Russia—was the failure of the bourgeoisie, as far back as 1848, to complete their middle-class political revolution and eliminate the structures of monarchy and aristocratic rule. But Kautsky offered no program to deal with this problem other than the rhetorical formula of revolution and the relentless march of history which would sharpen the crises of capitalism.

When in 1918 all three empires—Germany, Austria-Hungary, and Russia—collapsed, the socialist parties in these and other countries all adopted widely varying courses. Although capitalism by the turn of the century had become the predominant *economic* form of Western society, political structures and cultural traditions varied widely from one country to another. Paradoxical as it seems in Marxist terms, the cultural elements, more than the economic conditions in each country, account for the varying forms the socialist movement took. As Schumpeter remarked, every country had its own socialism.

*England.* In Germany a rigid class structure, reinforced by a militaristic code of honor, excluded the workers from society and led to a counter stiffness of doctrinal Marxist orthodoxy. In Great Britain, by contrast, the intelligence of an old gentry class, a deep tradition of liberty and of political rights, the lack of a militaristic tradition and even of a standing army, the long-established supremacy of Parliament over the monarchy, and the deep-rooted empirical conception of politics, which rendered ideological “all-or-none” terms distasteful—all made for a civil polity, one that accepted the existence of the socialist movement as a legitimate part of British society.

The British Labour party came into being in order to realize the rights of the working class in the society. The exemplar of peaceful—and piecemeal—social change, it was never Marxist, though it was based on strong class feeling and class loyalty. The sources of this position are threefold: There was, first, the deep and persistent strain of nonconformist Christian evangelism, which saw equality as a moral imperative. (The writings of R. H. Tawney, principally his books *Equality* [1931] and *The Acquisitive Society* [1920], which helped to shape the English socialist outlook, reflect this evangelism.) There was, second, the influence of the Fabians, principally Sidney and Beatrice Webb, who, representing one aspect of the Benthamite tradition, were social engineers for whom socialism was a tidy answer to the waste

and disorder of the capitalist world. And third, there was the large trade union movement, which saw in the Labour party a means of influencing the course of favorable legislation.

Whereas the Continental socialist parties participated actively in the affairs of the Second International, the British Labour party, before World War I, remained insular and rarely tried to impose its influence, as the Germans did, on other socialist parties. Much of this was due to the general isolation of the British from Continental social thought (one rarely encounters a discussion of “the state” in English political theory), which in turn has to do with the British temperament. As G. D. H. Cole, who knew Webb well, wrote:

Sidney Webb's first thought in dealing with any question he took up was to find an administratively workable solution; and apart from a very few essentially simple ideas—he did not trouble himself much about underlying philosophy. He was fully convinced that the trend of events in the modern world was towards Socialism, and that this trend would continue: so that he saw no need to put himself into revolutionary opposition to the main course of development. . . . He had what is sometimes called a “civil service” mind—that is, a habit of translating every idea into terms of the machinery needed to give it effect. . . . He was, however, impatient of dreamers, and uninterested in theories which he could not turn into practical schemes. (Cole 1953–1960, vol. 3, part 1, p. 210)

H. G. Wells summed up still another side of British socialism when he left the Fabian Society after a personal falling-out with Shaw and the Webbs. He caricatured all of them savagely in his novel *The New Machiavelli*. “Her soul was bony,” he wrote of the character named Altiora Bailey (Beatrice Webb). “If they [Altiora and her husband] had the universe in hand I know they would take down all the trees and put up stamped tin green shades and sunlight accumulators. Altiora thought trees hopelessly irregular and sea cliffs a great mistake.” The root image of their world, Wells wrote, was “an organized state as confident and powerful as modern science. . . . Individualism meant muddle, meant a crowd of separate undisciplined little people all obstinately doing things jarringly each one in his own way. . . . The organized state would end muddle forever” (Wells [1910] 1927, p. 193).

The Webbs spent most of their lives making detailed empirical inquiries into problems of social welfare and administration. While Marxism is usually associated with the idea of planning, Marx, as we have seen, never drew any specific blueprint of a planned society. Nor did the German Social Democrats, despite their large parliamentary rep-



resentation, even study industrial organization and indicate what they might do if they came into power. The Webbs and the Fabians, however, issued several hundred tracts providing detailed expositions of Labour thinking on both the local and the national levels. One can appreciate the thoroughgoing detail of Fabian research from a series of studies (conducted between 1898 and 1901) about the municipalization of different services—alcohol traffic, milk supply, pawnshops, slaughterhouses, bakeries, hospitals, and fire insurance (see Cole 1953–1960, vol. 3, part 1, pp. 215–216). In 1909, the Webbs's minority report on the operation of the poor laws set out in comprehensive detail the conception and policy of the welfare state (Great Britain . . . 1909). Later studies dealt with the general problems of the organization and control of industry. [See the biography of WEBB, SIDNEY AND BEATRICE.]

The Fabians operated primarily as an elite group and never sought a large membership. Their influence was felt through their published ideas, their research, and their propaganda. Before World War I, the Fabians were a constituent group in the Labour party; but in keeping with their avowed tactic of permeating all institutions of society that had the power to influence policy—the civil service, the professions, business groups, and local government—they drew upon all groups, nonsocialist as well as socialist, for help. After 1918, when the British Labour party adopted a new constitution and accepted as its basic program a Fabian policy statement drafted by Sidney Webb, relations between the Fabian Society and the Labour party became closer.

The Labour party has been unique among socialist parties, not only because of its open emphasis on "gradualism" but also because of its structure. Unlike the Continental socialist parties, based on individual membership, the British Labour party originated as a federation of unions and constituent socialist societies, and its funds were raised principally through levies on the union members. Until 1918 individuals could not become members directly. They became affiliated with the Labour party either through membership in the Independent Labour party, the Fabian Society, or the trade unions, and policy was worked out in negotiations between these organizations. Candidates were nominated by agreement between the unions and affiliates, on a roughly proportional basis. The individuals elected to Parliament then formed the parliamentary Labour party.

After 1918 the membership system was changed in order to organize local Labour parties directly

in each constituency, but the federated structure remained. The formulation of policy has therefore been a complicated affair. Since the trade unions have traditionally constituted the largest group in the party and vote en bloc in the Labour party conventions, the annual Trades Union Congress is an important arena for adopting resolutions. The annual Labour party conference of unions, affiliates (such as the consumer-cooperative movement), and local constituency parties sets policy. But this policy is only morally binding on the parliamentary Labour party. In office, the Labour party is responsible only to the parliamentary party, not to the Labour party as a whole. In this crucial respect, once again, the British Labour party is shaped by the structure of British politics and not by the conventional theories of socialist organization.

*France.* In England, the transition to a modern industrial society was accomplished peacefully by the economic, and to some extent social, blending of the rising plutocratic groups with the gentry, while a set of political compromises brought all sections of the country, including the working class, into the society. In Germany, the older feudal elements, using the state, had created powerful industries and maintained a political hegemony over the subordinated middle class. The working class, excluded socially, had built its own institutions, but paradoxically these very institutions had facilitated the integration of the workers into the society.

France lacked any such unifying features. Economically, it was a two-sector society with a large peasant and artisan class alongside a modern industrial economy. Politically the feudal structure had been broken, but the bourgeois parties were never able to establish their unambiguous control; the unstable balance of forces periodically opened the way to an adventurer attempting to seize power. The working class itself was split.

In Germany and England, the trade unions were part of the organized socialist movement because they hoped to achieve most of their aims through political concessions by the government rather than through direct economic bargaining. But in France, the trade unions were completely independent of the socialist parties. One wing, the syndicalists, based their gospel on Proudhon's anti-authoritarian and antipolitical ideas, and their anti-parliamentary bias on the betrayals of the 1848 revolution and the Commune. The French form of union organization, the Bourses de Travail, which stressed the local community of all trades rather than the nationwide organization of one industry or craft, expressed the tendency to create a new society of labor rather than to concentrate

on wages and working conditions. The other wing was that of "political socialism," but here, too, many of the temperamental weaknesses of French politics—its tendentiousness, its hyperbolic rhetoric, and its instability—were apparent in the French socialist movement.

In 1896, there were no fewer than six national socialist parties in France, each usually more interested in fighting the others than in fighting the opposition. By 1905, the six had been reduced to two national parties, one led by Jules Guéde, who was the spokesman for orthodox Marxism and whose following was chiefly in the industrial north, and the other by Jean Jaurès, a former professor of philosophy and a renowned orator—a humanist repelled by the aridities of Marxist dogma—whose following was among teachers, skilled workers, and intellectuals attracted to the idea of ethical socialism.

Under the pressure of the Socialist International, the two parties made an uneasy union, but the factions were still unable to agree on whether to enter coalition governments headed by bourgeois parties. The French parliamentary system, with its emphasis on multiparties, made it difficult for any single party to assume power. The art of government was the art of coalition. As socialist parliamentary strength rapidly increased, the socialists faced the problem of silently abstaining, thereby allowing rightist cabinets to govern, or entering center-left coalitions. Those who opposed coalitions argued that the assumption of governmental responsibility would weaken the militancy of the workers and would force the party to agree to non-socialist programs. Those who favored coalition, originally named the Possibilists, argued that in government, socialists could more easily defend the republic against reactionary forces—and at the same time help pave the way to socialism.

**Syndicalism.** If the British Labour party and Fabian gradualism together represent one end of the socialist continuum, revolutionary syndicalism, with its faith in direct action and the general strike, represents the other. The word "syndicalism" simply meant "unionism," but in the period preceding World War I it connoted an antiparliamentary, antireformist tendency deeply rooted in the Proudhon anarchist, antipolitical, antiauthoritarian tradition.

Revolutionary syndicalism was primarily a phenomenon of the Latin countries—France, Italy, and Spain—though syndicalist elements made a strong showing in the British labor movement among the seamen and the transport workers, and in the United States among the western miners

and loggers of the Industrial Workers of the World ("Wobblies"). Syndicalism never took hold in central Europe, the heartland of orthodox Marxism.

Marx and Engels had taught their followers to regard syndicalist tendencies as an expression of backwardness and immaturity, as a passing phase in the development of industrialism which would disappear after the emergence of the large-scale factory system and a modern industrial proletariat. Whatever the validity of that appraisal, syndicalism as it emerged in France was not only a despairing rebellion against industrial capitalism but, in its vision of the future, a protest against the destruction of free trade unionism under an authoritarian state socialism.

This aspect of syndicalism was formulated by Fernand Pelloutier, a journalist who had been active in the various Marxist movements in France. Disillusioned with the political parties, which were preoccupied with obtaining office and power, Pelloutier felt that the only protection workers could have against arbitrary managerial power—either in nationalized industries or in capitalist enterprises—would be workers' control of industry.

Syndicalism was important less as a doctrine for reorganizing society than as an attitude. It was hostile to parliamentary methods—and in 1905 the French trade union movement, in the famous Charter of Amiens, laid down the principle of strict independence from all political party involvement. Its injunction has been so strong that, unlike every other European trade union official, no French union secretary was allowed to take a parliamentary seat. The charter proclaimed the general strike as the instrument of revolution, a single collective action whereby the entire working class, by laying down its tools, could halt the operations of industry and in that "transforming moment" take power. Further, the charter glorified spontaneity rather than organization; and it emphasized the role of a conscious minority, an elite of revolutionary proletarians whose task it would be to lead an aroused working class into revolutionary action. [See *SYNDICALISM and the biography of SOREL.*]

**Guild socialism.** The syndicalist emphasis on workers' control has had recurrent appeal in working-class and radical movements. In Great Britain, where the "medievalist socialism" of William Morris and John Ruskin, firmly against industrialism and statism, caught on for a time, syndicalist ideas had a strange efflorescence before World War I in the "guild socialism" of A. R. Orage and G. D. H. Cole. [See *the biography of COLE, G. D. H.*]

The guild socialists, reacting against the administrative socialism of the Fabians, blueprinted a



decentralized socialist society in greater detail than any other socialist movement had. Politically, the guild state was to be a bicameral body—the one a geographical parliament based on local constituencies, the other a “functional” body made up of representatives of each trade or industry. The consumer, through Parliament, was to set the goals of production (e.g., the division between consumption and investment, the priorities of development); the Council of Guild Representatives, the producers, was to be responsible for the efficient management of industry. Each guild was to be a self-governing body, based on local councils, and was to set its own conditions of work. Each guild would receive money in proportion to its membership, but would pay wages in accordance with its own rules—either in equal shares or in differentials according to skill. Thus a national political and economic planning system was combined with the idea of cooperative workshops.

**Russian populism.** The course of world politics since 1917 has been dominated by the long shadow and the doctrinal pronouncements of Vladimir Il'ich Uli'ianov. Before 1914, however, Lenin played only a small role in the affairs of international socialism. He was a member, after 1905, of the bureau (executive committee) of the Second International, one of 69 persons representing 23 member countries. He was known personally to the leading figures of the socialist movement, but his works, not yet translated, were little known; and as an exile representing one of several fiercely quarrelsome sects, he carried little weight in the International.

Moreover, none of the prominent theoreticians of Marxism expected a socialist revolution in Russia: there was only a small industrial proletariat, and the country was still backward and feudal. In accordance with the theory of “necessary” stages of social development which Georgii Plekhanov (1883) had posited in founding the Marxist movement in Russia, this vast country still had to pass through the stage of capitalism, and the bourgeois middle-class democratic revolution was still to come. Once Russia could be led along the lines of Western social development, then political freedom, trade union freedom, and legal socialist activity would be achieved; after the democratic revolution, which was the role of the middle classes, would come the social revolution, in the more distant future. In effect, Russia was still “before 1848.”

Yet the revolution did occur and was shaped by certain peculiar features of Russian social history. Before socialism, the dominant radical tradition

in Russia had been populism, a doctrine associated in large measure with that remarkable exile Alexander Herzen. Herzen saw in the peasant communes the seeds of a future cooperative society that could bypass the harsh and disruptive effects of capitalism. From London, Herzen kept alive the liberal spirit of the Russian intelligentsia through his magazine *Kolokol* (The Bell), and in his home were to be found the major exiles from Russia.

The populism preached by Herzen idealized the peasantry and asserted, in almost mystic fashion, that the peasant was the source of wisdom and virtue. In the summer of 1873, roused by the appeal of Mikhail Bakunin, hundreds of students went to the countryside to “go to the people” and rouse them to action. Students disguised as workmen wandered the countryside, preaching revolution, but the peasantry, suspicious of their would-be saviors, simply turned them over to the police.

The episode was important in the history of populism, and its lesson was drawn most starkly by Peter Tkachev, one of the theorists of Russian populism. Insisting that the peasants as a mass—the people—were incapable of revolutionary creativeness and that only a “conscious minority”—the intelligentsia—could make the revolution, Tkachev sketched the kind of organization that would be necessary. It would have to be, he argued, a conspiratorial one, based on the principles of centralization of power and decentralization of functions. And it would have to be led by the intelligentsia. These two themes—the need for compact organization and the role of a revolutionary elite—were to bear fruit some twenty years later in the thinking of Lenin [see the *biography of LENIN*].

### Socialism between the world wars

The war that had begun in the summer of 1914 not only brought revolution to Russia; it signaled the collapse of international socialism. For several years the heat and lightning of war had flashed in Europe, and each time the international socialist movement had proclaimed its readiness to strike in order to prevent international conflagration. Its growing power seemed to assure a new foundation for the maintenance of peace. Yet in 1914, with very little dissent, the socialist parties of Germany, Austria, and France all voted to support their governments in the war. The German Social Democrats, the most powerful socialist party in the world, had in the past publicly dissociated themselves from the German state. And the Kaiser, in turn, had once called them “fellows without a country.” Now, with only one dissenting vote, the

parliamentary party gave full support to the budgetary war credits the government requested. There were tiresome quotations from Marx in support of the action: Marx had supported the principle of nationality; Marx had once proposed support of Germany in a war against Russia; and in 1891 Engels had said that in such a war Germany would be fighting for its national existence. Once scripture was being cited, the French party had its own rationalizations, heavily laden with quotations from Marx. So did the Austrians. The fact remained that when the crisis finally came, nationalism as an emotional idea proved to be stronger than class, and international solidarity proved to be a myth. The International was at an end.

**Polarization of belief.** The period between the two world wars saw Europe torn apart by the conflicting ideologies of communism and fascism. In the process, democracy and the socialist movement were the losers. Italy, Germany, Austria, and Spain came under fascist or authoritarian leadership. Even earlier, right-wing dictatorships took over Portugal, Hungary, and Rumania. Belgium and France were threatened by strong fascist movements. Only Great Britain and the Scandinavian countries were relatively free of these storms, though a small fascist group arose in England. The sociological reasons for these variations in fortune will be discussed in this section.

The October Revolution in Russia had brought before the European socialist movement the insurrectionary idea of the seizure of power. For more than forty years the idea had been an abstraction to the socialist movement. Lulled by the "inevitablism" preached by Engels and by the steady growth of electoral success, the socialist movements had assumed that at some distant time it might yet be necessary to seize, or at least to maintain, power that had been established legally; but no one took the idea seriously. Even inside Russia the idea, while fiercely debated, had an air of mimetic combat. Plekhanov had argued that men could act only insofar as social conditions allowed them to do so, i.e., only within the limits of the "laws" of history. But Lenin had apparently demonstrated the primacy of "will," at least within disorganized situations wherein a small group of determined men, acting skillfully and in disciplined fashion, could seize power.

Within Marxian theory there were actually three successive versions of the theory of taking power. The first, later presented in Lenin's "What Is to Be Done?" (1902), conceived of the proletariat as directed by a small group of professional revolutionaries drawn from the middle-class intelligent-

sia; the working class would support the middle-class revolution in the faith that there would be a second round wherein the proletariat—supported in Germany by a peasant revolt—could take over. But this conception passed with the end of the revolutionary wave of 1848–1850; and as industrialization and a new appraisal of the nature of factory life and the role of the proletariat emerged, a second version appeared. Now the emphasis was placed on the building of mass political parties led by workers who had achieved theoretical competence—typified in Germany by Bebel, who had been a carpenter, and in England by Keir Hardie, who was a miner. It was now felt that socialism need not come through insurrectionary tactics or coups led by small bands of professional revolutionaries, but peacefully, through parliamentary means or even simply in a show of strength.

After 1905 a few socialist theorists had argued that a new stage was emerging. These included Rosa Luxemburg, a Polish-Jewish intellectual with doctorates in philosophy and jurisprudence; Herman Gorter; Anton Pannekoek, a Dutch astronomer; and A. L. Helfand, a Russian-born economist who wrote under the name Parvus. Each of these socialists was influenced by developments in the Western labor movements rather than by events in Russia. In her book *The Accumulation of Capital* (1913), Luxemburg tried to extend Marx's economic doctrines by arguing that after a phase of imperialism, in which capitalists would seek to export capital surpluses, the capitalist system must inevitably break down and create a crisis. [See the biography of LUXEMBURG.] Gorter and Pannekoek had been close to anarcho-syndicalism, and Rosa Luxemburg and Parvus (in his early years) had been active in the left wing of the German social democratic movement. In one way or another, they all insisted that with the growing education of the working class, there would develop, during a crisis, a revolutionary spontaneity in the masses, and that to insist on party hierarchy and professional leadership would only lead to a dictatorship by the leaders.

**The Third International.** It was only after the Bolshevik victory in 1917, and the creation of the Third (Communist) International, that Lenin's earlier text was canonized in order to claim for the Bolsheviks a unique revolutionary knowledge and thus to enforce the hegemony of the Russian party over all other Communist parties.

Lenin summoned the revolutionary working-class groups to a conference, which met in Moscow on March 2, 1919, for the purpose of organizing the new International. Its main objectives



were the immediate seizure of power by working-class parties in Europe, the abandonment of false bourgeois democracy, and the establishment of a dictatorship of the working class for the systematic suppression and expropriation of the exploiting classes.

There was an apocalyptic fervor in the air. World revolution seemed palpably near. Shortly after the first congress of the Communist International, a Soviet republic was proclaimed by Béla Kun in Hungary, and another in Bavaria by left-wing socialists. It seemed as if the only thing needed to carry out a successful revolution was steely revolutionary will. A second world congress of the Comintern (the shorthand name for Communist International) was convoked in Moscow in July 1920. It was no longer a small gathering; delegations came from parties in a dozen countries.

The chief feature of the meeting, which gave organizational shape to the international communist movement, was the drafting of 21 points as conditions for membership in the Comintern. The purpose of these points was to create in each country a disciplined, conspiratorial party whose chief purpose would be to combat the old socialist leaderships and to assert the binding, from-the-top-down, authority of the Comintern over each national party. Throughout Europe and in the United States, such socialist leaders as Ramsay MacDonald in England, Kautsky and Rudolf Hilferding in Germany, Morris Hillquit in the United States, and Jean Longuet in France had opposed participation in the war and had taken a "middle" position against the reformists. But in the opinion of the new Comintern, these leaders had to be rejected and exposed as much as those of the right wing, and had to be fought just as bitterly. The 21 points commanded communists to split every socialist party and trade union in the world, to organize an underground machine in addition to the public activities of the party, to disorganize as much as possible the army of each country, and to reject any cooperation with "social patriots and middle-group people."

By 1923 the revolutionary tide had receded all over Europe. The communists had completely misread both the character of the labor movements of western Europe and the social structure of those societies. For a short period the communists engaged in adventurism and even in putschism: the Red Army marched into Poland to advance the revolution, only to be defeated by Pilsudski, once a nationalistic socialist, who a few years later set up an authoritarian regime; insurrections were

planned, in 1923, in Saxony and Thuringia; and an abortive uprising in Hamburg failed. But after 1923 it was clear that, for the time being at least, Europe had achieved some measure of political and economic stabilization. The Soviet Union itself turned, under Lenin, to the problem of what to do with power in a single country. The large Norwegian Labor party, as well as such syndicalist leaders as Jack Tanner in England and Alfred Rosmer and Pierre Monatte in France, withdrew from the Comintern because of the centralization of party structure. What was left in Europe was the wreckage of the socialist movement in half a dozen countries and the fear of revolution that drove the middle classes to support right-wing groups. Within the Comintern, the hegemony of the Russian party was complete, and within a short time the International itself became principally an arm of Soviet foreign policy, rather than an independent instrument of revolution. [See COMMUNISM, *article on THE INTERNATIONAL MOVEMENT.*]

The sharp turn to the left in the Soviet Union in 1929, and Stalin's effort to consolidate his rule by turning on his erstwhile right-wing political allies, coincided with a world-wide economic depression and the rise of fascism in Germany and other countries. For the communists these events heralded the final crisis of capitalism, and they awaited a fresh wave of revolutionary activity. After an analysis of fascism in Italy, communist theorists argued that fascism was the last stage of monopoly capitalism; and since it could not solve the inherent contradictions of the capitalist crisis, inevitably the revolution was again at hand. From this analysis the communists concluded that the chief obstacle to their victory was not the capitalists but the socialists, who still "misled" the majority of the working class. In several instances, the communists even worked with the Nazis in order to diminish socialist influence. They voted with the Nazis in the Prussian Landtag to bring down the Social Democratic government. They cooperated with the Nazis in the Berlin street-car strike in 1932 in order to increase disorder in Germany.

**Socialists in government and opposition.** The communists thus became the implacable enemies of the democratic regimes in central and western Europe; and the middle classes in many countries, principally Italy and Germany, out of fear of the left and because of economic crises, often voted for the extreme right. However, an important element contributing to the weakness of the democratic regimes was the inability of the socialists, owing to the contradictory attitude toward capital-

ism and democracy inspired in them by Marxian dogmatics, to provide any effective leadership or support for democratic societies.

In 1931 the reconstructed Socialist International consisted of parties with more than six million dues-paying members. The total parliamentary vote for socialist candidates was almost 26 million. More than 1,300 socialist deputies sat in the parliaments of their countries. Some 360 daily newspapers spoke for the labor movement. Yet, remarkably, this large force was almost completely paralyzed when the crises occurred.

The root problem was an old one. The socialist movement, true to its Marxist heritage, did not believe that capitalist society could be reformed. When the socialists, particularly of the right wing, were thrust into office because of the failure or the unwillingness of any other party to rule, they followed the most orthodox of economic policies, because "the crisis has to run its course." Believing, from a Marxist point of view, that the reason for the depression was a disproportion in growth between the producer goods sector and consumer goods sector, they tried to use up the resultant "overproduction" so that a better proportion between producers' and consumers' purchasing power would emerge, leading to an upswing.

As Adolf Sturmthal has pointed out in *The Tragedy of European Labor* (1943), the socialist movement, with all its strength, was basically a pressure group seeking social concessions from the state for the immediate benefit of the working class. But it had neither an economic program nor any clear idea of planning. State intervention arose out of unorthodox economic theories, such as John Maynard Keynes's in England and Gunnar Myrdal's in Sweden, or the unorthodox financial policies of Hjalmar Schacht, who had been made president of the Reichsbank by the German socialists when Hitler began the rearmament that revived the German economy. Nowhere except in Sweden, and later in the planning ideas of Hendrik de Man, did the socialists have any idea of what to do about the depression [see the biography of MAN].

*Italy.* The socialists, in a different way from the communists, also misread the nature of fascism. For example, fascism was barely mentioned in the major report of the 1928 International Socialist Congress on the political situation in Europe. It was seen as an idiosyncrasy of the Italians; and its ideology, emotional roots, and irrational quality were not understood. Yet Italy did foreshadow quite clearly the fate of the other nations in central Europe.

Shortly after the war, Italy seemed on the verge of a proletarian revolution. In the general election of 1919 the socialists won two million out of a total of 5.5 million votes, and the leadership of the party had passed into the hands of the left wing, which openly asserted that the next step would be the "creation of a Socialist Republic and the establishment of a proletarian dictatorship." Workers had spontaneously begun to seize factories, and the peasants of Sicily and the south had appropriated the uncultivated holdings of absentee landlords. As Sturmthal put it: "Continuous unrest, strikes, factory occupations, expropriation of land—all this convinced the middle class that a revolution was impending and that the democratic middle-class state was powerless to stave off the danger. Public opinion became more and more convinced that a strong man was needed to establish law and order" ([1943] 1951, p. 182).

The decisive, dramatic incident occurred in August 1920, when a wage dispute in the metal industries led to sudden "stay-in" strikes in which 500,000 workers occupied the factories, kept the machinery going, and assembled arms to resist evacuation. Workers in other industries called on their leaders to order the taking over of other factories. But the socialist leadership, divided and uncertain, hesitated; and finally a pact was reached with the industrialists whereby the employers agreed in principle to the union's demand for workers' control of production. This was the high point of the revolutionary tide, and then a new force appeared, the Fascists.

Organized by Benito Mussolini, a former leader of the left wing of the Italian Socialist party, the Fascists preached anticapitalism, nationalism, and the necessity of violence. With his *squadristi*, Mussolini went into the streets to break up working-class meetings and to beat up working-class leaders. In 1921 an effort was made to form a socialist-liberal coalition government and save the country from the threatened civil war. The right-wing socialists made the proposal, but the idea was vetoed by the left wing. By 1922 a form of civil war had spread in Italy. In the large urban industrial centers of the north, strongholds of the socialist movement, the city administrations passed into the hands of Mussolini's *squadristi* through terror and intimidation. Bologna, Genoa, Livorno, Milan, and finally Naples were taken over by the fascists. A general strike called by the trade unions on August 31, 1922, failed ignominiously, despite the united support of the labor movement; and middle-class opinion swung even more strongly to the



fascists. At the invitation of the king, and with the support of the army, bureaucracy, and big business, Mussolini was invited to become premier. For two years he ruled by parliamentary means, but after the assassination of Giacomo Matteotti in June 1924 and left-wing withdrawal from the Chamber of Deputies as a moral protest, Mussolini became more openly dictatorial: trade unions, political parties, and cultural organizations were either disbanded or placed under fascist control, and local autonomy was abolished. By 1926 parliamentary government had vanished. [See FASCISM.]

*Germany.* In November 1918, the Germany of Wilhelm was no more. The Kaiser had abdicated, and Friedrich Ebert, a former saddlemaker who was now the head of the Social Democratic party, installed himself as head of the new republic. But the socialists themselves were split into three factions. The "majority socialists" represented the right wing of the party. The "independent socialists," led by Kautsky and Bernstein (together for the first time in twenty years), had opposed the war and now favored a radical program of economic reform. The extreme left, led by Luxemburg and Liebknecht, knew that the German republic was going to be a middle-class state and wanted to organize a proletarian party prepared for an eventual revolutionary opportunity. But younger socialists took over the left wing en masse, overruled Luxemburg and Liebknecht, and began to prepare the "Spartacus group," as the left wing called itself, for immediate revolution.

In this situation the attitude of the majority socialists was decisive. The Workmen's and Soldiers' Councils, which had sprung up spontaneously on the Soviet model, elected the majority socialists to the leadership of the new Council of People's Commissars. But the majority socialists feared a repetition of the Russian chaos and sought first to achieve stability with the cooperation of some of the military.

When the Spartacus group initiated a rebellion in early 1919, it was put down by Gustav Noske—the majority socialist appointed as commander-in-chief of the army—with the help of the Free Corps, created and led by former imperial officers. After the uprising was quashed, Free Corps officers cold-bloodedly murdered Liebknecht and Luxemburg, who had loyally supported their comrades despite their opposition to the venture. When, in the spring of 1919, left-wing socialists—and later the communists—took over the newly proclaimed Bavarian republic, the Free Corps was used to take Munich and to murder hundreds of the insurgents,

thus seriously reducing the authority and prestige of the socialists. In March 1920, Reichswehr troops led by a little-known nationalist named Wolfgang Kapp mutinied and marched into Berlin. The republican government could not muster enough loyal troops to defend the capital, and fled to Stuttgart. A powerful general strike, joined by all the factions, defeated the putsch in four days. When Kapp was routed from Berlin, communist-led workers in the Ruhr tried to continue the general strike. The newly formed Weimar coalition—majority socialists, the Catholic Center, and Conservative Democrats—sent the Reichswehr into the Ruhr to crush the revolt.

In the end the majority socialists, as well as the republic, were the losers. In the general elections of June 1920, the majority socialists and the middle-class parties of the Weimar coalition lost heavily, and the nationalist parties gained. A new cabinet consisting of the Catholic Center and the right-wing German People's party took office. It was clear that on the right as well as on the left the republic itself had only shaky support among the German people.

The socialists were given one more chance. In May 1928 the German Socialist party, now reunited because the independents refused to accept the *Diktat* of the Comintern, emerged as the strongest party in the Reichstag. Although they lacked an absolute majority, the socialists took office, with Herman Müller as chancellor and Hilferding, the famed socialist theoretician, as minister of finance. But a year later Germany, along with the rest of the Western world, was plunged into the depression, and the socialists had no economic policy to meet the crisis. Hilferding, mindful of the ruinous inflation of the early 1920s, followed an orthodox deflationary policy which reduced purchasing power and increased unemployment. The strength of the labor movement defeated the employers' efforts to reduce wages and salaries; the fault, however, lay not with the employers, who had to reduce production costs or get out of business, but with the state, which had failed to work out any active policy. Instead of tapping idle capital funds, the government worked above all to balance the budget, or at least to reduce budget deficits, even if this meant reducing unemployment insurance benefits. Several German socialist economists favored devaluation or the abandonment of the gold standard, a monetary policy later associated with Keynes. But they were opposed, on the ground that this would lead to economic and political nationalism. In all essentials the socialists followed

a policy of laissez-faire: the depression had to "run its natural course." After all, as any Marxist knew, capitalism could not be reformed.

*England.* A similar dilemma confronted the British Labour party. In 1918 it had adopted a socialist program for the first time; the new social system was described as a thing that would emerge gradually out of capitalism, by a series of piecemeal changes. The Labour party was then still weak, third in size after the Conservatives and the Liberals. Five years later, following a prolonged period of unemployment which the Conservatives had been unable to cope with, the Labour party emerged as the strongest English party and, supported by the Liberals, in January 1924 formed the first Labour government in British history. The government carried out some modest social reforms, but its tenure was short. When the British Foreign Office published the so-called Zinoviev letter, a set of instructions from the head of the Comintern to British communists on antimilitarist tactics—a letter now conceded to be a forgery—the electorate voted strongly Tory, and the Labour government was ousted.

The trade unions, disappointed by their failure in politics, turned to more militant economic action. The coal miners, always the most militant, had a genuine grievance—their wages had recently been cut. (The problem was one of government monetary policy; in 1925 England returned to the gold standard at the prewar pound-dollar exchange rate, and the prices of British exports were above the world market level.) The miners, refusing to accept the wage cut and demanding the nationalization of industry, went on strike in May 1926. With the support of the entire trade union movement, this strike soon widened into a general strike, the first in English history. Railwaymen, local transport workers, builders, printers, iron and steel workers, all walked out, almost completely paralyzing London and other parts of Great Britain. The strike had had no revolutionary aim—its only purpose had been to support the miners—but when the government stood firm, the unions, uncertain of their next step, retreated. After nine days the general strike was called off, and its main result was that the left wing lost influence and the right wing gained complete control of the labor movement.

In June 1929 the British Labour party had its second chance. In the general election of that year, the party won 287 of the 615 seats in the House of Commons and, with the support of the Liberals, formed the second Labour government, with Ram-

say MacDonald as prime minister. But the problem that soon confronted the German socialists was already bedeviling England. Though other countries at the time were still enjoying prosperity, England, because it could not compete in world markets, had a great deal of unemployment. The Labour government was pledged to make far-reaching social reforms; the business community demanded reduced taxes and a retrenchment in social policy. A collision was inevitable. One way out would have been devaluation or strict exchange controls to keep gold from leaving England. Either course would have been an acknowledgment of the end of Britain's domination of the international economy—which she had maintained for almost a hundred years—and a new policy of economic nationalism. This the Labour government refused to do.

When the depression hit England full force, the Labour government had its back to the wall. The flight of capital from London had become a flood, and the Bank of England warned that unless the budget was balanced, the pound would fall. In orthodox fashion, the Labour government was committed to free trade and to defense of the gold parity of the pound. MacDonald proposed, as an economy measure, to cut unemployment benefits; and when the trade union elements in the party rejected such a cut, or any reduction in social services, MacDonald split the Labour government and, taking 14 colleagues with him, formed a national coalition with the Conservatives and the Liberals. The national government itself failed to stem the tide, and in September 1931 Britain went off the gold standard, introduced protectionism and a tariff, and began, under a Tory government, to set up economic dikes in an effort to save itself from the floods of world-wide depression. The Labour party, though it still retained some strength among the electorate, suffered a great loss of parliamentary strength—from 287 to 52. For nine years it sat in opposition until it joined the Churchill government of national unity in 1940. In 1945, for the first time in its history, the Labour party won a clear electoral majority.

*Austria.* In February 1934, after four days of bloody fighting, the reactionary regime of Engelbert Dollfuss destroyed the Austrian social democracy. This was a blow that struck the international socialist movement especially hard, for Austrian social democracy had been the model for all proud socialists. It was a disciplined party and had the support of almost all of the working class. Its leaders and theoreticians, particularly Friedrich Adler and



Otto Bauer, had been respected for their courage and their contributions to socialist thought. The party, while revolutionary in its aims, maintained a sanity and realism in political affairs which had prevented it from imitating the adventurism of the Hungarian communists in 1919 or the feckless policy of the German Social Democrats. In 1927, at the peak of its strength, the Austrian Socialist party polled 42 per cent of the total electoral vote, and in Vienna the socialists won a majority of almost two-thirds. Half a million of Vienna's two million inhabitants were dues-paying party members, and the city was a showcase of municipal achievement. In 1919 the socialists had joined with the Christian Social party in a coalition that gave the country stability and permitted sections of the old Austro-Hungarian Empire to achieve their independence. The first president of the Austrian republic, Karl Seitz, was a socialist, as was its first chancellor, the noted legal scholar Karl Renner.

The problem of the Austrian socialists was twofold: their strength was entirely in the urban areas, while the surrounding countryside and the middle classes supported the Christian Social party; and Austria was under direct and steady pressure from Italy to crush the socialist movement and establish a fascist regime. When the Nazi vote began to increase in Austria, as a result of Hitler's prestige and his direct aid to the new Nazi party of Austria, Dollfuss, a Catholic, was faced with the choice of coming to terms either with the Nazis or with the socialists. With the support of Mussolini (who wanted an independent Austria as a buffer against Germany), Dollfuss suppressed both the Nazis and the socialists. But the defeat of the socialists meant that the Austrian government had lost its main hope of independence. Four years later Kurt von Schuschnigg, who succeeded Dollfuss when the latter was assassinated by the Nazis, invited the underground socialists and the trade unions to support him against the Nazis. However, while negotiations were going on, Nazi troops invaded Austria and the country was annexed by Germany.

*France.* Only in France, among the major countries of western Europe, was a fascist threat—that of Colonel de la Rocque and the Cagoulaards, the right-wing group supported by the army—beaten back. In June 1936, a socialist government headed by Léon Blum took office. It was the first time that the party officially entered a coalition government, but it did so—also for the first time—with the support of the communists, who had abandoned their cry of "social fascism" and now proclaimed the need for a popular front to resist

fascism. Within a year a French New Deal had been inaugurated. The trade unions, never before recognized by French employers, now were granted collective bargaining rights. A social insurance system was established for the first time. Through public works and wage increases, Blum tried to increase the purchasing power of the workers and, thus, to restore prosperity. But, curiously, Blum resisted the idea of economic planning, fearing that it was essentially "statist" and fascist in character. Caught between the rejection of capital exchange controls by its political allies, the Radicals, and the communist opposition to devaluing the franc, the first Popular Front government fell.

*Sweden.* Only in Sweden did the socialists have any real success, and they achieved it by abandoning the orthodox economic policies that the German Social Democratic and British Labour governments had followed. In 1932 a Swedish labor government was formed for the first time. It decided that balancing the budget on a year-to-year basis made little sense and that the government itself had to intervene in the economy. The government embarked on a set of extensive public works and financed these endeavors not by taxes, which simply would have shifted the existing purchasing power, but by borrowing money from idle capital funds. Public investment during a depression had to be expanded to compensate for reduced private spending. By following a steady policy of economic expansion, the Swedish government managed to eliminate unemployment by 1938. Five years earlier it had been as high as 164,000. These "Swedish stepping stones to full employment," as A. P. Lerner described the process (1944), wholly unorthodox at the time, have become commonplace economic practice in almost all Western countries.

The rest of the melancholy story of European socialism before 1939—the Spanish Civil War, the Soviet purge trials and the Hitler-Stalin pact—while crucial to the history of Europe, is less relevant to the discussion of socialist theory and doctrine.

## Socialism since World War II

*Revisionism and peasant revolution.* In the theory and practice of socialism after 1945 there were two completely unexpected developments. One was the rise of "revisionism" in the communist countries and movements of Europe. That is, the highly centralized command economies were modified by the trend toward a market and profit system, dogmatic ideology was eroded and tended to be replaced by pragmatic and instrumental pol-

icies, and force and violence were discouraged as a means of fostering revolutionary change in non-communist countries. In fact, with the extension of planning and public control in the noncommunist countries of Europe as a parallel to the growing decentralization of planning in communist countries, some theorists have argued that a "convergence" was taking place between the communist social systems and those of the West.

The second development was the rise of the peasantry as a revolutionary force—or at least as a presumed revolutionary force—in colonial countries, bolstered by a new theory that the revolutions in the last third of the twentieth century would be made not by the proletariat, which had lost its élan through *embourgeoisement*, but by the peasantry. For the peasants were now, in the words of the "Internationale," the wretched of the earth. Marx, in many of his writings, had scorned the peasantry as intrinsically reactionary and small-minded because of its preoccupation with private property; he used the phrase "rural idiocy" more than once. But modern methods of communication and a network of trained political cadres have since the 1920s been able to weld the peasantry into an active political force. The Chinese Communist party under Mao Tse-tung, following the destruction of its urban base of support when Chiang Kai-shek smashed the Shanghai commune in 1927, had reorganized itself as a peasant party and had won power by enlisting peasant support. Fidel Castro, though a middle-class intellectual, made his revolutionary appeal through the peasantry, while the Cuban working class, including the communist-dominated trade unions, "coexisted" tacitly with the dictator Batista. In Algeria, in South America, in Vietnam, the peasantry, not the urban working class, became the focus of revolutionary appeal.

The spread of socialism as word and doctrine, especially in the years after World War II, presented itself as a paradox, particularly in terms of the original intention and predictions of Marx. It was assumed by Marxists that the triumph of socialism would occur first in the industrialized countries, as a result of the contradictions and crises of the capitalist economic system; but this triumph has taken place primarily in backward countries and in agrarian societies. Marxism was an ideology of protest *against* the course of industrialization, but it has become, instead, the ideology of industrialization; not a creed of social justice, welfare, and the equitable distribution of products, but a rationale for centralized controls, postponement of consumption, and rapid economic growth.

Since the Soviet Union embodied these developments, it became an important model for many of the new states that were seeking to embark on the difficult road of industrialization. The Stalinist regime had apparently succeeded in transforming Russia from an agrarian country into an industrial one. Despite enormous waste during collectivization, it had achieved high rates of economic growth, and its successes were most dramatically symbolized by the technological achievement of being the first nation successfully to launch man-made vehicles into outer space.

**Socialism in western Europe.** In the West, the socialist movement saw the complete triumph of what, in classical doctrine, would be called reformism. During the war against fascism, almost all the socialist parties had joined the governments of national unity, either at home or in exile. The sense of the nation and democracy took priority over the ideas of class and capitalism. In the Western countries, the labor and socialist parties had become completely of their societies, as well as *in* them.

One can identify five features, common to almost all the socialist parties of western Europe, which marked this new practice and doctrine:

(1) The complete abandonment of the idea of revolutionary methods and violence as a means to power; the complete acceptance of parliamentary means; and the complete readiness to participate in nonsocialist coalition governments. The anguished theoretical debates which had earlier split the French, Belgian, Austrian, and other parties on the question of entering "bourgeois governments" had vanished.

(2) The transformation of the socialist and labor parties from *class* parties, speaking only for working-class interests, to *people's* parties seeking a more inclusive concept of general welfare. In some instances, as in Germany, there was the realization that as a working-class party the Social Democrats would be consigned to being a permanent minority; in other instances, as in England, there was the recognition that the changing class structure of society, particularly the rise of a technical and salaried middle class, made it imperative for the Labour party to speak for these social groups as well as for the working class.

(3) The recognition that the definition of socialism as a social and economic ideal was inseparable from the idea of democracy, both as a means and as an end. The Marxist concept that lingered through the 1930s—that democracy was a "bourgeois" concept and only a mask for class rule—was rejected. By democracy as a means, the socialist parties meant the full guarantee of the rights



of free speech and free assembly, and the maintenance of political rights for the opposition. As an end, democracy was defined as the free consent of the governed.

(4) The surrender of the idea of nationalization or state ownership of the means of production as a "first principle" of socialism, and the substitution of public control of enterprise and planning as the means of achieving economic growth and equitable incomes. The sectarian orthodoxy that within a capitalist state one could not plan for social ends was replaced by the theory that governmental powers could be used for the gradual transformation of the economy and that a "mixed economy" of public and private enterprise was the most desirable solution.

(5) A complete opposition to totalitarianism. While the socialist movements had long been in ideological opposition to communism, what the new attitude meant, in practice, was support of the military and political aspects of the Atlantic Alliance against the Soviet bloc and an identification with the national interests of each Western country, through NATO, against Russian expansionism. Nothing, perhaps, more strongly revealed the degree of change than the choice of Paul-Henri Spaak, once the firebrand leader of the Belgian left, as secretary-general of NATO.

*Germany's Social Democrats.* Most of the transformed socialist attitudes were expressed at the Frankfurt Congress of 1951, which symbolized the revival of the Socialist International. Although no longer the powerful voice of a confident movement or the platform for thunderous international pronouncements, the Socialist International still provided some doctrinal identification for those who still called themselves socialists. Its Declaration of Principles, unanimously adopted, stated the new common conceptions of socialism almost a hundred years after the initial statements of the First International, as formulated in Marx's inaugural address of 1864.

"Socialism," the declaration reads, "aims to liberate the people from dependence on a minority which owns or controls the means of production. . . . It aims to put economic power in the hands of the people as a whole and to create a community in which free men work together as equals" (Lowenthal 1951, p. 113).

Private versus public control remains the decisive contrast, but private control is not identified with private ownership, nor public control with state ownership; and the phrase "exploited classes" has been superseded by "the people." Democratic planning becomes the basic means for the creation

of socialism, and public ownership is regarded as only one of a number of different means, to be used where necessary. "Democratic socialism," says the statement, "therefore stands in sharp contradiction both to capitalist planning and to every form of totalitarian planning; these exclude public control of production and a fair distribution of its results" (*ibid.*).

Democracy, in this revised credo, is not only a means to the achievement of socialism—a "favorable battleground for the class struggle," as the older view had it—but an integral aspect of socialism itself. "Accordingly," as Richard Lowenthal put it, "the conflict between democratic socialism and communism no longer appears as a disagreement about means to a common end, but as a conflict of fundamental ends between the adherents of a democratically controlled economy and the defenders of the despotism of a managerial bureaucracy" (*ibid.*).

Nowhere is the change more striking than in the new program of principles adopted in 1959 by the German Social Democratic party in Bad Godesberg. The German SPD had long been a reformist party, but its very adherence to outworn dogmas, as we have seen in its attitude toward the possibility of reforming capitalism in the 1920s and 1930s, crippled its ability to act realistically in economic and political affairs. But the Bad Godesberg program rejects the very idea of Marxism. The name of Karl Marx and the concept of Marxism are missing from the declaration of principles, and terms like "class" and "class struggle" are carefully avoided.

In modern political analysis, party programs are rarely taken seriously. Usually they are formal documents, paying lip service to expected pieties that historically made the parties' activities legitimate. But, as F. R. Allemann points out:

The German social-democracy is a "programme party" of the first water: in its history of nearly a hundred years it has always attached the greatest importance to basing its policy on a solid system of principles laid down in a fixed programme rather than merely on the urgent issues of the moment. In this respect at least it has remained true to its Marxist tradition: in its view any political action that is intended to influence and transform society must be based on an analysis of that society. (1960, p. 67)

This was the view that guided the formulation of the *Erfurt Program* in 1891; it is the view behind the Bad Godesberg declaration of 1959.

The overriding idea in the Bad Godesberg program is that a modern industrial society cannot be ordered by a single uniform principle. Common ownership is recognized as a legitimate form of

public control, but the analysis centers not on property but on economic power, which must be subject to public control. For the first time a socialist party states explicitly that private ownership of the means of production is entitled to "protection and promotion" so long as it does not hinder the construction of an equitable social order. The idea of subjecting the entire economy to central planning is rejected, and the party accepts the free market, wherever there is real competition. The basic formula is "as much competition as possible—as much planning as necessary," and the SPD not only describes free consumer choice and the free choice of place of employment as "all-important foundations" of freedom, but lauds free competition and entrepreneurial initiative as important elements of Social Democratic economic policy. The economic pattern envisaged is that of a "mixed economy" in which the private profit motive has a place and in which the control of great economic power is the central task of a libertarian economic policy.

In line with the lessons of the previous twenty years, particularly those of the Soviet experience, the party declared that the problem of economic power cannot be settled simply by transferring power from private hands into that of the state. Being dependent on an uncontrolled state bureaucracy is not necessarily better than being subject to private capitalist power. The conclusions, however, are rather vague: common ownership is to be "organized according to the principles of self-government and decentralization," and in the managerial structures "the interests of the workers and employees must be represented just as much as the public interest and that of the consumers" (*ibid.*).

The aim of the new socialist program is a plural society, and this idea extends not only into economics but into the cultural sphere as well. Thus the party refrains from proclaiming any "ultimate truths" and states that neither a state nor a political party should have any power in religious and philosophical spheres. As Allemann says, "Marxism set out both to 'interpret' and to 'change' the world." The German Social Democrats, having rejected the myths behind the ideology of socialism, no longer "claim to provide a universally valid philosophy nor do they believe any longer that their policy is in accord with the irrevocable laws of social dynamics" (*ibid.*).

*Britain's Labour party.* The British Labour party never had German socialism's attachment to world philosophy, and its sense of crisis was less severe. The wellsprings of evangelical feeling, ex-

pressed in a strong commitment to equality and social justice, still provide the emotional justification of socialism for the English Labourites; but a different problem has confronted the leadership. Socialism, as it was accepted in the West, had been a singularly distributivist doctrine. It presumed that capitalism had solved all the problems of production and that its failures had to do with social injustice and inequality. But the problem of economic growth was found to persist in all social systems. How does one step up the rate of output and increase production without overt controls and coercion, and without inflation?

In 1945 the British Labour party, for the first time in its history, won a majority in Parliament and assumed sole charge of the government. For six years it ruled, and during that time it laid down the permanent foundations of a welfare state in England. Social services were extended, and for the first time a comprehensive system of medical care was established. A number of basic industries, principally coal, railways, transport, and steel, were nationalized; but, to the dismay of the government, nationalization did not provide any automatic answers to the problem of growth. Coal and railways had been sick industries, but the Labour policies, although they helped maintain full employment, did not materially increase the productivity of these industries. In 1951 Labour was voted out of office and replaced by a Tory government. Its defeat caused the British Labour party to ponder its program. In office it had exhausted the "intellectual capital" inherited from the early Fabians. What would it do if it was voted into power again? The major task of intellectual renovation was attempted by C. A. R. Crosland in his book *The Future of Socialism* (1956), which became the working text for Hugh Gaitskell, the new leader of the party. Crosland's main argument was that the Labour party must give up the shibboleth of nationalization and, in economic matters, promote the modernization of industry in whatever social form it could best be done, whether private or public. The main content of socialism, Crosland said, is not economic but social: to seek an equitable division of wealth by reducing the role of inherited property as the means of achieving a privileged place in the society; and to erase social class distinctions, by a major reform of the educational system. Only by releasing talents which remain unfulfilled, because of the social class system, could Britain eventually find new vigor in its industry and its culture.

For the traditionalists in the party, both left and right, the Crosland-Gaitskell vision smacked of



heresy. An effort on Gaitskell's part to eliminate clause four—which pledges the party to nationalization of industry as a major doctrinal tenet—from the British Labour party constitution, was rebuffed. Gaitskell died before the British Labour party was again voted into power in 1965, but the revisionism he espoused had clearly taken hold in the party.

*Italy's Socialist party.* The one socialist party in the West which maintained a traditional left-wing line was the Italian. Forged in exile and deeply conscious of the intense class heritage of Italian workers, the Italian Socialist party, led by Pietro Nenni, maintained an electoral alliance with the communists for more than ten years, and still vaguely preached the class struggle. A small group of moderate Socialists, led by Giuseppe Saragat, split away from the Italian Socialist party in 1947 to join the coalition government led by the Christian Democrats.

The Khrushchev disclosures in 1956 of Stalin's misdeeds, the Soviet intervention in Hungary in the same year, the general disillusionment with communism, and the growing feeling in Italy that no effective role could be played without entering a coalition government eventually led Nenni, in 1963, to join a center-left coalition and finally, in 1966, the two wings of Italian socialism were reunited.

Whatever the socialists in Italy, or in any of the Western countries, could offer, it was no longer the old certitudes of doctrinaire Marxism. Few of the parties had any sure answers about how to build a new social order. All that remained was a lingering vision—and hope.

*Socialism in the "third world."* To speak of the *tiers monde* as a sociological entity is to group together geographical areas that are vastly different in their historical and class structures and similar only in their poverty, their instability, and, in most cases, their quest for modernization.

The area encompassed by north Africa and the Middle East, from Morocco to Egypt and Syria, was the cradle of ancient civilizations, an area ceaselessly marauded by invading armies for thousands of years. Its populations have absorbed a great jumble of genetic strains, and in the last several hundred years the mixture of Arab, Turkish, and European (French, English, and Italian) cultures has blended into a cosmopolitan *mélange*. It is a world with a small, westernized, intellectual elite, a strong, traditional religious structure, and a vast mass of poor peasants. Southeast Asia has an array of old societies and ancient empires which, in most instances, were held together after the departure

of the former colonial rulers by a thin administrative class, trained by these rulers, while the political parties or military regimes learned how to govern. Africa south of the Sahara is a motley collection of tribal societies grouped in political areas that were carved out by the old imperial powers. These areas were almost never formed on a basis of indigenous unity, and there has been no educated or technical class available for the many complex tasks of government and economic development.

The social structures of these areas are vastly different from those of Europe, and it is difficult to see where Marxist categories could be applied. In south and southeast Asia, primordial relations of blood, religion, language, or tribe are expressed as communal attachments; and nations have been split by ancient ethnic, tribal, caste, regional, and religious differences. India has Muslim-Hindu divisions, caste lines, secularist-traditionalist splits, and ten major linguistic divisions. Burma, since gaining its independence, has been plagued by insurrections of the Shan, Karen, and Kachin peoples, split by linguistic, religious, tribal, and regional differences. Ceylon's Sinhalese and Tamil communities differ in language, race, caste, and religion. Indonesia has the classic division between the central island, Java, and the outer islands, as well as a secularist-Islamic controversy. The Malay states are split between the Chinese and the Malays, and the disunity is compounded by indigenous tribal groups in Sarawak and Sabah.

In Africa, within the predominant tribal societies, there has been little class differentiation, since most of the land has been held collectively and life has been led communally. But between tribes there has been grisly hostility, and the rectification of borders between the new nations provides grist for conflict that will go on for generations. In the Islamic world, in countries as divergent as Morocco and Pakistan, theocratic structures are fused with the political and administrative apparatus, and this fusion in turn shapes a distinctive social structure.

The major tasks which most of these countries have set for themselves are modernization and industrialization. How far modernization necessarily must upset the traditional and religious values of a society is a question that divides sociologists. But it is clear that in many of these countries the new elites are seeking to upset or supplant the traditional values, often because the source of their own authority or of their routes to power rests on different criteria. In these countries socialism is often seen as the means of creating new, modern-

ized states. Since the end of World War II, Arab socialism, African socialism, and Asian socialism have emerged as new ideological flags and new socioeconomic forms.

The socialism of these new elites is vastly different from the traditional visions of the nineteenth-century socialist prophets, intellectuals whose ideologies were universalistic and humanistic. The new ideologies are parochial, instrumental, and largely the creation of political leaders. While the driving forces of the old ideologies were social equality and political freedom, economic development and national power inspire the new ideologies.

Socialism appeals to the new elites in a number of ways: as a doctrine, it identifies them with a historical-progressive movement; as an ideology, it seeks to create an identity that can transcend tribal and communal boundaries; and as a social system, it concentrates power in the hands of the elites, providing for the nationalization of the basic industries of the countries and their direction under some ministerial control. Although the socialism of the West has come to favor decentralization and a mixed economy, the socialism of the new states focuses on central planning, state ownership and direction of enterprises, and one-party regimes. Socialism, in effect, serves as a means of mobilizing the society for industrial and social transformation. [See MODERNIZATION.]

*The Middle East.* No socialist movement of the Middle East calls itself, as the Western movements did, the representative of the working class. These movements, not only in fact but often in theory, are based on an alliance of the new middle class (largely salaried state employees), the peasants, and the workers, with the middle class openly taking the lead. To the extent that a theory of representation exists, it is based on an undifferentiated general will. In practice, the effort is being made to give nationalism a social content.

This is clearly the case with the socialism of Gamal Abdel Nasser, in the United Arab Republic. Before he came to power in 1952, there had been no large-scale socialist movement in Egypt. (The Egyptian Socialist party was in fact fascist.) Nor had Nasser and the army group he represented ever espoused socialist ideas. But once he was in power, his proclaimed "Six Objectives of the Revolution," plus his opposition to foreign business, led him increasingly to speak of his socialist faith. In 1954 he declared: "We consider that the state has tutelage over both private and public property and the responsibility for the protection of the individual against all economic and social exploi-

tation" (Halpern 1963, p. 244). In 1956, after the abortive Anglo-French-Israeli invasion of Egypt, Nasser appropriated the Suez Canal and foreign banks, insurance companies, and export firms. By 1962 he had nationalized all large Egyptian enterprises, expanded land reform, and placed all important sectors of the economy under the control of socialist planners. Democracy, the Egyptians were told, would come only gradually.

Egyptian socialism is not based on any mass or cadre party. Rather, it is being installed from the top down, through the initiative of a small technocratic elite whose support is largely military. Whether it can survive its first, charismatic leader is a moot point, because, lacking a social base, Egyptian socialism has not been able to institutionalize its reforms.

The other major socialist movement in the Middle East is the Arab Socialist Resurrectionist party, commonly called the Ba'ath party. Organized in Syria in 1953 as a merger of two smaller parties, it declares itself "a national, populist, revolutionary movement fighting for Arab unity, freedom and socialism." Its nationalism is not restricted to any particular Arab state, but to the Arab people as a whole, and it is against "all other denominational, factional, parochial, tribal or regional loyalties." Its socialism, in the words of its programmatic statement, seeks the "guarantee [of] the continuous growth of the nation in its spiritual and material development; and it will guarantee close brotherliness among its individual members" (*ibid.*, p. 240). Full of rhetoric about revolution and struggle, the Ba'ath party has said little about the concrete steps it would take to achieve socialism. Like many such movements, it is seeking to develop a new faith to supplant the traditional creeds. Islam is never mentioned by name; the implication, as Halpern points out, is that it belongs to an earlier era. In practice, the Ba'ath party, caught in the nationalist rivalries between Egypt, Syria, and Iraq, has split accordingly.

*North Africa.* In north Africa the socialist creed has been fashioned largely by French-educated native intellectuals, and it assumed a revolutionary coloration, particularly in the late 1950s, through the vigorous movements for independence from France. In Algeria, where the struggle was most bitter, the FLN (National Liberation Front) was composed mainly of Algerian trade unionists (most of whom worked in France and contributed money to the struggle), middle-class elements, and Berber nomads. After Algeria became independent, Ahmed Ben Bella, surrounded by various Marxist and Trotskyite advisers, sought to introduce



workers' self-management in industry and to collectivize Algerian farms. But as a result of the flight of French technicians and agronomists, Algeria was soon in severe economic trouble; and Ben Bella's efforts to solidify his personal power led to a coup by his ally, Colonel Houari Boumedienne, who controlled the army. Boumedienne pledged to continue "Algerian socialism," but his program was the typical drab mixture of statist enterprises and private landholdings characteristic of so many other new states. For the French left, ironically, Algeria had in the meantime become the emotional symbol of a "new awakening" of the underdeveloped world. Frantz Fanon, a Negro psychiatrist born in the French West Indies and educated in France, renounced his French citizenship during the Algerian civil war and wrote a new Communist manifesto, *The Wretched of the Earth* (1961), which preached the necessity of violence as purification. Fanon also praised the noble qualities of the peasantry, still uncorrupted by bourgeois values—unlike the European working class [see the biography of FANON]. For Jean-Paul Sartre and other leftist intellectuals, the Algerian war was a crisis of conscience in which they declared their civil disobedience against the French state. Yet the revolution ended in tawdry factionalism and a gray, repressive regime.

A quieter path to socialism was taken in Tunisia in the 1960s by the Neo-Destour party, later renamed the Socialist-Destour party. Led by Habib Bourguiba, a French-educated lawyer, Tunisia became independent in 1956, without the fierce violence of Algeria. A one-party regime was installed, and it proceeded to nationalize the major enterprises of Tunisia and to develop a planned economy. The Socialist-Destour movement is actually a middle-class party, and the trade unions are firmly subordinated to the state. Bourguiba, as the charismatic leader, has imposed a vigorous cult of personality on Tunisian society—numerous streets, monuments, enterprises, and other public activities are named for him. It is still a "first-generation" country, and it remains to be seen whether it can institutionalize its new reforms and maintain a stable political structure without undue group conflict when Bourguiba is succeeded.

*Sub-Saharan Africa.* The 1960s have seen the mushrooming of a new doctrine, largely unknown even five years earlier, which African leaders call African socialism. The meaning of the phrase, however, is very difficult to pin down (Friedland & Rosberg 1964). Few African socialists have any idea what they mean concretely by socialism. President Léopold-Sédar Senghor of Senegal, for ex-

ample, who has tried to provide a framework in his book *African Socialism* (1959), sees socialist humanism as a combination of Marx and Engels with Teilhard de Chardin, in which the Jesuit paleontologist-philosopher's ideas of "corpusculization" and "complexity" (the effects of outer and inner gravity in creating an increasing complexity of matter) are joined with the dialectic to create "the general law of complexity-consciousness." Whether it is the quasi mysticism of Senghor, trying to create a new definition of man, or the tough-sounding Marxism-Leninism of President Sékou Touré of Guinea, seeking to join the idea of a cadre party with the "communaucratic" values of precolonial African society, the idea of African socialism is one of pure rhetoric—a language to impress, to inspire, to intimidate—rather than a set of guidelines for specific action. Its usefulness is hortatory rather than instrumental.

What "African socialism" means to these countries is the effort to be distinctive and modern, to assume some special character in the contemporary world, and, in a more prosaic way, to deal with the problems of social identity, economic development, and new class formations in the emerging societies.

The problem facing most of the new states in Africa—31 black African nations gained their independence in the decade between 1955 and 1965—is to find some unifying symbols beyond parochial and local identities in order to instill a sense of pride and action in their people. If many citizens of these new states have little feeling of nationality, they have nonetheless found some kinship through the idea of being African. African socialism has been a new myth, something to take the place of the anticolonial passions that fueled many of the independence movements. Seeking a specific content for African socialism, its ideologists have stressed the indigenous character of the communal ownership of land, the egalitarian character of the society, and the extensive network of social obligations which ties clans together. Thus, it is argued, capitalism and private property are "unnatural" to Africa.

But though the indigenous elements of traditional society have been present in Africa, there still remains the crucial question whether such institutions can carry over into the complex, modern industrial society that these countries want to build. By and large, most African countries committed to the idea of socialism see the government as exercising the primary responsibility for accumulating capital, directing investment, and building an infrastructure for the society. The im-

mediate practical difficulty is that Africa has had no large entrepreneurial class of its own from which modern talent can be drawn. The only large non-European commercial and manufacturing class, particularly in east Africa, has consisted of Indians; and because of their clannishness and their ties to the mother country, they are regarded as unassimilable or as aliens. Nor do most African countries, except perhaps Nigeria and a few others that were under British rule, have a sufficiently large number of administrators, engineers, or managers. Another dilemma is that the capital accumulation for most African countries is dependent on earnings from agricultural and other primary products, and these are subject to sharp fluctuations on the world market or are dependent on the good will of the advanced industrial countries; and good will is a notoriously frail foundation on which to build an economic policy.

Many countries seek a way out by industrialization. But rapid industrialization, if at all possible, involves a large-scale transformation of the farming class (which in most African countries is between 80 and 90 per cent of the population), new class distinctions, and new social group tensions, either between different interest groups or between populations that have become urban and cosmopolitan and those that have remained local and parochial.

Senghor, when he turns to practical matters, says that Senegal must proceed slowly. As he wrote in *African Socialism*: "We have rejected prefabricated models . . . we have observed that formulas like 'priority for heavy industry,' or 'agrarian reform' have no magic power within themselves; applied dogmatically, they have produced partial failures. That is why we established priorities as follows: 'infrastructure, rural economy, processing industry, heavy industry,' in line with reasonable requirements and our realities" (1959, p. 157). In keeping with this ordering principle, Senegal has earmarked a substantial part of its budget for "social development," by which it means health and hygiene, municipal administration and housing, and education. The economy as a whole is not socialist, but mixed: communal in agriculture, a public-private combination in utilities, and private in banks, commerce, and industry.

Under Kwame Nkrumah, Ghana set out vigorously to create a socialist organization of the economy. While private capital was invited, particularly for the construction of the Volta Dam, state enterprises predominated in banking, manufacturing, construction, fishing, forestry, and electric power. But the ambitious schemes ran far ahead of the

reality. The trade unions protested the squeeze on wages. Capital reserves and foreign exchange were squandered. Grandiose and costly projects—presidential residences and palaces, as well as the large-scale Pan-African political headquarters—were built at enormous expense. One-party rule became more and more authoritarian: opponents were jailed or murdered, the courts were reorganized to stifle their independence, and the freedom of the university was abridged, while Nkrumah's decrees became increasingly personal. In 1966, while Nkrumah was traveling abroad, a revolt by the army ended his regime.

Most African countries have found that their major problems are primarily political: to maintain that fundamental stability which will allow any planning, economic or otherwise, to proceed. The initial glittering promises made by the first generation of political leaders have not been realized. Within the first decade of independence, military coups racked half of the new nations, and in the others rigged elections and strict bans on opposition have kept the initial regimes in power. For these countries the tensions of social change, and the need to create institutional mechanisms to deal with them, constitute the major problem confronting the societies. A Kenya White Paper of 1965 on African socialism and its application to planning in Kenya, said soberly that socialism, even the socialism of a welfare state, would be a long time in coming. The immediate need was to transform the economy from a subsistence to a market economy; to develop land and to introduce modern agricultural methods. Nationalization, since it does not always lead to additional resources for the economy as a whole, would be used only when other means of control were ineffective. The commitment is to socialism as an ideal, not as an ideology (Harris 1965).

*South and southeast Asia.* In January 1953, representatives of a dozen Asian socialist parties met in Rangoon for the first Asian Socialist Conference. The principal Asian figures—Jayaprakash Narayan and Asoka Mehta of India, Ba Swe and Kyaw Nien of Burma, Sutan Sjahrir of Indonesia—were internationally renowned. Leading Western figures came as fraternal delegates: Clement Attlee, the former Labour prime minister of England, headed the delegation from the Socialist International, and Milovan Djilas the one from Yugoslavia.

At this time, the governments of half a dozen countries—India, Burma, Ceylon, Indonesia, Cambodia, and Singapore—described themselves as socialist, for socialism was the predominant ideology of the area. It seemed that socialism in Asia



was making a triumphal entrance on the world historical scene.

Some dozen years later, Narayan had joined the voluntary *bhoo dan* movement of the Gandhian Vinoba Bhave, and Asoka Mehta had left the Socialist party; Ba Swe and Kyaw Nien were in jail; and Sjahrir, having spent six years in prison, died shortly after his release in 1967. The Asian Socialist Conference itself was no more. Of the socialist governments, India had increasing economic and political trouble; Burma had come under a military dictatorship; Ceylon had swung back to a nonsocialist government; Indonesia was racked by an abortive Communist-inspired revolution in 1966, and by a military-led counterrevolution which massacred several hundred thousand communists and finally stripped Sukarno, the first president of the regime, of all his powers; Cambodia was perched precariously between East and West, struggling to maintain its independence during the Vietnam war; Singapore, after first joining Malaya to form the new state of Malaysia, broke away and remained a small independent enclave still proclaiming itself socialist. What had happened to these Asian socialist parties, and what had happened to the socialist governments?

One anomaly was that in these countries, which proclaimed themselves officially socialist, the Socialist parties were rarely in full control of the governments. In Burma, the Burmese Socialist party was for many years part of the ruling coalition of the Anti-Fascist People's Freedom League, and Ba Swe had at one time been premier of his country. In Indonesia, Sjahrir had been its first prime minister, in 1945, heading a coalition government. But the Burmese Socialist party was abolished in 1962, when General Ne Win seized power and proclaimed Burma a socialist state; and the Indonesian Socialist party, wiped out at the polls in 1955, was abolished by presidential decree in 1960, when Sukarno proclaimed *Resopim* (Revolution, Indonesian Socialism, and National Guidance) as the official ideology. Except for the period of the independence movements and the struggle against the colonial powers, and for a few short years after each country won its independence, the socialist parties of Asia have not played a major role in their nations, despite the official adherence of the governments to a socialist ideology. The regimes called themselves "socialist" but were not based on the socialist parties.

Unlike the socialist movements of the Middle East and Africa, which scarcely existed before World War II, the socialist parties of Asia had a long history of involvement with the international

socialist movement and socialist thought. Jayaprakash Narayan, who became a revolutionary Marxist at the University of Wisconsin in the late 1920s, founded the Bihar Socialist party in 1931, was acting general secretary of the Congress party during the civil disobedience movement in 1932, and founded the Congress Socialist party in 1933, first as a wing of the Congress party and after 1939 as an independent party. The party organ in the early 1930s, the *Congress Socialist*, was edited by Asoka Mehta. Like the socialist parties of Europe, the Congress Socialist party had its difficulties with the communists, who stigmatized it as being "social fascist" and a "left maneuver of the bourgeoisie." Its experiences with the communists led the Indian Socialist party in later years to take a strong anti-Soviet, antitotalitarian position. Sjahrir had been a leader of the Perhimpunan Indonesia (PI), the student association of Indonesians studying in Holland in the 1920s, and on his return to Indonesia in 1932, he was interned by the Dutch and kept in prison for ten years; upon his release, he organized and directed underground resistance against the Japanese.

The Asian socialist and student movements had long had "tutelary" relations with the European socialist parties, particularly since most of the countries were under imperial rule; and freedom for the colonies had been an important plank in the programs of the British Labour party, the Dutch Labor party, and others. Thus the leaders of the Asian socialist parties had gone through the doctrinal vicissitudes of the European socialist movements, and in most instances had themselves adopted democratic socialist positions. The weakness of these parties, particularly the Indonesian, was that they were primarily parties of intellectuals, with some following among the workers but no influence or base among the peasantry in societies that were overwhelmingly agrarian. Their orientation was largely urban, and as Marxists they regarded the industrialization of their countries as the normal path of development. Moreover, as parties of intellectuals, with only a small hold on the trade unions, they were peculiarly subject to the factionalism and divisiveness of intellectual groups who lack an anchorage. Individual socialists, such as Jayaprakash Narayan, Asoka Mehta, or Sutan Sjahrir, were at times enormously influential in their countries—but only as individuals and because of their individual talents, not as party leaders.

Most south Asian countries called themselves socialist even when the leading parties, such as the Congress in India or the Nationalist in Indo-

nesia, had no doctrinal faith, simply because they assumed that state direction of the economy and state planning were necessary for economic growth. And the difficulty for most of these countries was that, lacking any experience in planning, lacking the vital managerial talents, and, in many cases, lacking the needed resources, they wasted available resources and made mistakes that could not be absorbed by the economies. Many of the Asian countries, mesmerized by the idea of industrialization, assumed, in the early 1950s, that, following the Russian model, this meant a priority for heavy industry. Burma, for example, decided to build a steel mill, even though its main "natural resource" was the huge amount of gun scrap left behind by the invading Japanese and the defending British troops after World War II. Other expensive show-cases were also built, and precious capital and foreign exchange were wasted on large airports and other "visible" features put up to impress foreigners with the progress and modernity of the new country.

It would be a simplification, of course, to attribute the failures and setbacks primarily to defects in economic planning. As was indicated earlier, southeast Asia, more than almost any part of the world, is plagued by communal conflicts, so that the regimes have been torn by internal conflicts and by threatened and actual insurrection. Also, these countries have been caught in the crossfire of Great Power rivalries, particularly the demand of the United States for participation in SEATO (the Southeast Asia Treaty Organization) and the diplomatic and military pressure of the Chinese, who seek to extend their influence in southeast Asia. But apart from the real political and sociological problems, a crucial fact was that each of these countries plunged ahead on a course called socialist which not only was failing in practice but also was increasingly contradictory to the historic socialist ideals. The latter problem was foremost in the thinking of Jayaprakash Narayan, perhaps the outstanding figure of Asian socialism in the mid-twentieth century. Speaking at the second—and last—conference of the Asian socialist parties, in 1956, Narayan declared:

All our countries, except Japan, are backward economically and many are desperately poor. Naturally, therefore, our attention goes first of all to the problem of economic growth. There is nothing wrong in that, but the mischief starts when we begin measuring "socialist achievements" in terms of tons of steel and kilowatts of electricity. Economic growth, even rapid economic growth, is known to have occurred both under Capitalism and Fascism. Mere economic development is not a measure of socialism. I do not wish

to suggest that it is not the business of Socialists to see that more wealth is produced. What I wish to emphasize is the danger of equating socialism with economic development, and of sacrificing the values of socialism at the altar of that development. . . .

The main, if not the whole, emphasis is still being placed on the control and use of the power of the State. Everywhere socialists are organized in political parties which are attempting to seize power and hoping thereafter to build a new society. . . . But as I have said before the ideals of socialism remain far in the distance. The reason seems to me to be a wrong approach to these ideals. All of us agree that socialism is a way of life, an attitude of mind, a certain ethical behaviour. What is not so universally recognized is that such a way of life, attitude and behaviour cannot be imposed from above by dictates of the Government or by merely nationalizing industry and abolishing capitalism. Construction of a socialist society is fundamentally construction of a new type of human being . . . if human reconstruction is the key to socialist reconstruction, and if that is beyond the scope of the State, the emphasis in the socialist movement must change from political action to such work of reconstruction. (Quoted in Rose 1959, pp. 258-259)

Contemporary sociology and mature Marxism share a bias against utopian thought. What the Marxists called objective conditions, sociologists call structural constraints. The idea is the same: that the range of alternatives open to any society is limited by the starting point, the kind of resources, the degree of differentiation, and the like. Certain chosen paths impose certain imperatives: industrialization requires the creation of a technical class and a new kind of educational system; the increasing structural differentiation of a complex society requires a kind of social coordination different from a simple, top-down command system. Necessarily, such a perspective tends after a while to bring out the more limited choices of means of action, and in the process the ends are often forgotten or become ritualized. Socialism as a belief has been both a system of means and a system of ends; and throughout its history the means have invariably tended to diminish the ends.

In his youthful writings, Marx was most vividly concerned with the ends of human action. In the *Economic and Philosophic Manuscripts of 1844*, a halfway house in the development of his thought, Marx speculated on what the phase immediately following the socialist revolution might be. The negation of private property, he said, is not the aim of socialism, for what it does is to usher in a period of what he called raw or crude communism. Crude communism does not transcend private property but universalizes it; it does not overcome greed but generalizes it; it does not



abolish labor but extends it to all men. The aim of socialism, as a fully developed naturalism, as humanism, is to go beyond communism, beyond necessity, and therefore beyond history—which itself is a form of determinism—to a world which resolves the “strife between existence and essence, between objectification and self-confirmation . . . and between the individual and the species.” It is a world in which a human being no longer feels “divided” or alienated from what he believes his essence as a social being, as a person free to make his own future, can be (Marx 1844, pp. 104–114 in the 1919 edition).

This is the permanent utopian—and even religious—component of socialism, a quest, as ancient as man's fall from grace, to unify himself with an ultimate and to find a world of freedom. It remains a world beyond.

DANIEL BELL

[See also COMMUNISM; ECONOMIC THOUGHT, article on SOCIALIST THOUGHT; MARXISM; MARXIST SOCIOLOGY. Related doctrines and ideas are discussed in ANARCHISM; CAPITALISM; NATIONALISM; PLURALISM; SYNDICALISM; UTOPIANISM. Other relevant material may be found in IDEOLOGY; PLANNING, ECONOMIC; PLANNING, SOCIAL; SOCIAL MOVEMENTS; STRATIFICATION, SOCIAL; WELFARE STATE.]

#### BIBLIOGRAPHY

- ADLER, MAX (1910) 1923 *Der Sozialismus und die Intellektuellen*. 4th ed., rev. Vienna: Wiener Volksbuchhandlung.
- ALLEMANN, F. R. 1960 Farewell to Marx. *Encounter* 14, no. 3: 67–69.
- BEBEL, AUGUST (1883) 1910 *Women and Socialism*. New York: Socialist Literature. → First published in German.
- BERNSTEIN, EDUARD (1899) 1909 *Evolutionary Socialism: A Criticism and Affirmation*. London: Independent Labour Party. → First published as *Die Voraussetzungen des Sozialismus und die Aufgaben der Sozialdemokratie*. A paperback edition was published in 1963 by Schocken.
- BORKENAU, FRANZ (1938) 1962 *World Communism: A History of the Communist International*. Ann Arbor: Univ. of Michigan Press. → First published as *The Communist International*.
- BRAUNTHAL, JULIUS 1961–1963 *Geschichte der Internationale*. 2 vols. Hanover (Germany): Dietz.
- BUBER, MARTIN 1947 (1950) *Paths in Utopia*. New York: Macmillan. → First published in Hebrew.
- CHALMERS, DOUGLAS A. 1964 *The Social Democratic Party of Germany: From Working Class Movement to Modern Political Party*. Yale Studies in Political Science, No. 10. New Haven: Yale Univ. Press.
- COLE, G. D. H. 1953–1960 *A History of Socialist Thought*. 5 vols. New York: St. Martin's; London: Macmillan. → Volume 1: *Socialist Thought: The Forerunners, 1789–1850*, 1953. Volume 2: *Marxism and Anarchism, 1850–1890*, 1954. Volume 3: *The Second International, 1889–1914*, 2 parts, 1956. Volume 4: *Communism and Social Democracy, 1914–1931*, 2 parts, 1958. Volume 5: *Socialism and Fascism, 1931–1939*, 1960.
- CROSLAND, C. A. R. (1956) 1957 *The Future of Socialism*. New York: Macmillan.
- DOLLÉANS, ÉDOUARD; and CROZIER, MICHEL (editors) 1950–1959 *Mouvements ouvriers et socialistes: Chronologie et bibliographie*. 5 vols. Paris: Éditions Ouvrières. → The most recent general bibliography of socialist publications and works about socialism; promises to become the standard guide.
- DURBIN, EVAN F. M. (1940) 1965 *The Politics of Democratic Socialism: An Essay on Social Policy*. London: Routledge.
- DURKHEIM, ÉMILE (1928) 1958 *Socialism and Saint-Simon (Le Socialisme)*. Yellow Springs, Ohio: Antioch. → First published in French. A paperback edition was published in 1962 by Collier.
- ENGELS, FRIEDRICH (1845) 1958 *The Condition of the Working Class in England*. Oxford: Blackwell. → First published in German.
- ENGELS, FRIEDRICH (1878) 1959 *Anti-Dühring: Herr Eugen Dühring's Revolution in Science*. Moscow: Foreign Languages Publishing House. → First published as “Herrn Eugen Dühring's Umwälzung der Wissenschaft” in a series of articles in *Vorwärts* (Leipzig).
- ENGELS, FRIEDRICH (1880) 1935 *Socialism: Utopian and Scientific*. New York: International Publishers. → First published in French.
- Fabian Essays. 6th ed. (1889) 1962 London: Allen & Unwin. → First published as *Fabian Essays in Socialism*; contributors include George Bernard Shaw and Sidney Webb.
- FANON, FRANTZ (1961) 1965 *The Wretched of the Earth*. Preface by Jean-Paul Sartre. New York: Grove. → First published in French.
- FRIEDLAND, WILLIAM H.; and ROSENBERG, CARL G. JR. (editors) 1964 *African Socialism*. Stanford Univ. Press.
- GAY, PETER 1952 *The Dilemma of Democratic Socialism: Eduard Bernstein's Challenge to Marx*. New York: Columbia Univ. Press. → A paperback edition was published in 1962 by Collier.
- GEIGER, THEODOR (1948) 1949 *Die Klassengesellschaft im Schmelztiegel*. Cologne (Germany): Kiepenheuer. → First published in Danish.
- GRAY, ALEXANDER 1946 *The Socialist Tradition: Moses to Lenin*. London: Longmans.
- GREAT BRITAIN, ROYAL COMMISSION ON POOR LAWS AND RELIEF OF DISTRESS 1909 *The Minority Report of the Poor Law Commission*. Edited by Sidney and Beatrice Webb. 2 vols. London: National Committee to Promote the Break-up of the Poor Law. → Part 1: *The Break-up of the Poor Law*. Part 2: *The Public Organisation of the Labour Market*.
- HALÉVY, ÉLIE (1938) 1965 *The Era of Tyrannies: Essays on Socialism and War*. Translated by R. K. Webb. Garden City, N.Y.: Doubleday. → First published in French.
- HALÉVY, ÉLIE 1948 *Histoire du socialisme européen*. Paris: Gallimard. → Published posthumously.
- HALPERN, MANFRED 1963 *The Politics of Social Change in the Middle East and North Africa*. Princeton Univ. Press.
- HARDIE, J. KEIR 1907 *From Serfdom to Socialism*. London: Allen.

- HARRIS, BELLE 1965 *Growth and the Good Relationship: Kenya's White Paper on African Socialism*. *New African* 4:142-144.
- HULSE, JAMES W. 1964 *The Forming of the Communist International*. Stanford Univ. Press.
- JABAR, KAMEL 1966 *The Arab Ba'ath Socialist Party*. Syracuse Univ. Press.
- JAURÈS, JEAN L. (1894-1914) 1964 *L'esprit du socialisme: Six études et discours*. Paris: Gonthier Éditions.
- JAURÈS, JEAN L. (1906) 1908 *Studies in Socialism*. 2d ed. London: Independent Labour Party. → First published in French.
- JOLL, JAMES 1955 *The Second International, 1889-1914*. London: Weidenfeld & Nicolson. → A paperback edition was published in 1966 by Harper.
- KAUTSKY, KARL (1892) 1910 *The Class Struggle (Erfurt Program)*. Chicago: Kerr. → First published in German.
- KAUTSKY, KARL (1902) 1916 *The Social Revolution*. Chicago: Kerr. → First published in German.
- LANDAUER, CARL 1959 *European Socialism: A History of Ideas and Movements From the Industrial Revolution to Hitler's Seizure of Power*. 2 vols. Berkeley: Univ. of California Press. → Volume 1: *From the Industrial Revolution to the First World War and Its Aftermath*. Volume 2: *The Socialist Struggle Against Capitalism and Totalitarianism*.
- LE BON, GUSTAVE (1898) 1965 *Psychology of Socialism*. New York: Fraser. → First published in French.
- LENIN, VLADIMIR I. (1902) 1961 *What Is to Be Done?* Volume 5, pages 347-529 in Vladimir I. Lenin, *Collected Works*. 4th ed. Moscow: Foreign Languages Publishing House. → First published in Russian.
- LENIN, VLADIMIR I. (1916) 1964 *Imperialism, the Highest Stage of Capitalism: A Popular Outline*. Volume 22, pages 185-304 in Vladimir I. Lenin, *Collected Works*. 4th ed. Moscow: Progress. → First published in Russian.
- LENIN, VLADIMIR I. (1917) 1964 *The State and Revolution: The Marxist Theory of the State and the Tasks of the Proletariat in the Revolution*. Volume 25, pages 381-492 in Vladimir I. Lenin, *Collected Works*. 4th ed. Moscow: Progress. → First published in Russian.
- LERNER, ABBA P. 1944 *The Economics of Control: Principles of Welfare Economics*. New York: Macmillan.
- [LÉVI, PAUL] (1936) 1950 *Histoire du socialisme en France*, by Paul Louis [pseud.]. 5th ed. Paris: Rivière.
- LICHTHEIM, GEORGE 1966 *Marxism in Modern France*. New York: Columbia Univ. Press.
- LOWENTHAL, RICHARD 1951 *The Principles of Western Socialism*. *Twentieth Century* 150:112-119.
- LUXEMBURG, ROSA (1904-1922) 1961 *The Russian Revolution: Leninism or Marxism? With an introduction by Bertram D. Wolfe*. Ann Arbor: Univ. of Michigan Press. → First published in German.
- LUXEMBURG, ROSA (1913) 1964 *The Accumulation of Capital*. New York: Monthly Review. → First published in German.
- MCKENZIE, ROBERT T. (1955) 1963 *British Political Parties: The Distribution of Power Within the Conservative and Labour Parties*. 2d ed. New York: St. Martins.
- MAN, HENDRIK DE (1926) 1928 *Psychology of Socialism*. London: Allen & Unwin. → First published in German.
- MANUEL, FRANK E. 1956 *The New World of Saint-Simon*. Cambridge, Mass.: Harvard Univ. Press.
- MANUEL, FRANK E. 1962 *The Prophets of Paris*. Cambridge, Mass.: Harvard Univ. Press.
- MARKHAM, FELIX 1952 *Introduction*. In Saint-Simon, *Selected Writings*. Edited and translated by Felix Markham. Oxford: Blackwell. → A paperback edition was published in 1964 by Harper.
- MARTOV, IULII O. 1938 *The State and the Socialist Revolution*. New York: International Review.
- MARX, KARL (1844) 1964 *Economic and Philosophic Manuscripts of 1844*. New York: International Publishers; London: Lawrence & Wishart. → First published posthumously in German in 1932. Sometimes referred to as the "Paris Manuscripts of 1844."
- MARX, KARL (1844-1850) 1926 *Selected Essays*. Translated by H. J. Stenning. London: Parsons.
- MARX, KARL (1867-1879) 1925-1926 *Capital: A Critique of Political Economy*. 3 vols. Chicago: Kerr. → Volume 1: *The Process of Capitalist Production*. Volume 2: *The Process of Circulation of Capital*. Volume 3: *The Process of Capitalist Production as a Whole*. Volume 1 was published in 1867. The manuscripts of volumes 2 and 3 were written between 1867 and 1879. They were first published posthumously in German in 1885 and 1894.
- MARX, KARL; and ENGELS, FRIEDRICH (1845-1846) 1939 *The German Ideology*. Parts 1 and 3. With an introduction by R. Pascal. New York: International Publishers. → Written in 1845-1846, the full text was first published in 1932 as *Die deutsche Ideologie* and was republished by Dietz in 1953. A paperback edition was published in 1965 by International Publishers.
- MARX, KARL; and ENGELS, FRIEDRICH (1848) 1963 *The Communist Manifesto*. New York: Russell.
- MARX, KARL; and ENGELS, FRIEDRICH (1848-1898) 1962 *Selected Works*. 2 vols. Moscow: Foreign Languages Publishing House. → Includes works published posthumously.
- MARX, KARL; and ENGELS, FRIEDRICH (1875-1891) 1938 *Critique of the Gotha Programme*. New York: International Publishers. → Written by Marx in 1875 as "Randglossen zum Programm der deutschen Arbeiterpartei." First published with notes by Engels in 1891.
- MEHTA, ASOKA 1959 *Studies in Asian Socialism*. Bombay: Bharatiya Vidya Bhavan.
- MICHELIS, ROBERT (1911) 1959 *Political Parties: A Sociological Study of the Oligarchical Tendencies of Modern Democracy*. New York: Dover. → First published as *Zur Soziologie des Parteiwesens in der modernen Demokratie*. A paperback edition was published in 1966 by Collier.
- MILL, JOHN STUART (1848) 1961 *Principles of Political Economy, With Some of Their Applications to Social Philosophy*. 7th ed. Edited by W. J. Ashley. New York: Kelley.
- MORGAN, ROGER P. 1965 *The German Social Democrats and the First International 1864-1872*. Cambridge Univ. Press.
- PELLING, HENRY 1954 *The Origins of the Labour Party 1880-1900*. New York: St. Martins; London: Macmillan.
- PELLING, HENRY 1961 *A Short History of the Labour Party*. New York: St. Martins; London: Macmillan.
- PLEKHANOV, GEORGI V. (1883) 1959 *Sotsializm i politicheskaia bor'ba (Socialism and the Political Struggle)*. Moscow: Gospolizdat.
- RIZZI, BRUNO 1939 *La bureaucratisation du monde; Le collectivisme bureaucratique: Quo vadis America?* Paris: Privately published.
- ROSE, SAUL 1959 *Socialism in Southern Asia*. Oxford Univ. Press.



ROSENBERG, ARTHUR (1932) 1939 *A History of Bolshevism: From Marx to the First Five Years' Plan*. London and New York: Oxford Univ. Press.

ROSENBERG, ARTHUR (1938) 1939 *Democracy and Socialism: A Contribution to the Political History of the Past 150 Years*. New York: Knopf.

ROTH, GUENTHER 1963 *The Social Democrats in Imperial Germany: A Study in Working-class Isolation and National Integration*. Totowa, N.J.: Bedminster Press.

SAINT-SIMON *Selected Writings*. Edited and translated with an introduction by Felix Markham. Oxford: Blackwell, 1952. → A paperback edition was published in 1964 by Harper.

SCHORSKE, CARL E. 1955 *German Social Democracy, 1905-1917: The Development of the Great Schism*. Cambridge, Mass.: Harvard Univ. Press.

SCHUMPETER, JOSEPH A. (1942) 1950 *Capitalism, Socialism, and Democracy*. 3d ed. New York: Harper; London: Allen & Unwin. → A paperback edition was published in 1962 by Harper.

SENGHOR, LÉOPOLD-SÉDAR 1959 *African Socialism: A Report to the Constitutive Congress of the Party of African Federation*. New York: American Society of African Culture.

SENGHOR, LÉOPOLD-SÉDAR 1964 *On African Socialism*. New York: Praeger.

SOMBART, WERNER (1896) 1909 *Socialism and the Social Movement*. London: Dent; New York: Dutton. → First published as *Sozialismus und soziale Bewegung im 19. Jahrhundert*. The title was later changed to *Sozialismus und soziale Bewegung*. The 10th edition was published in 1924 as *Der proletarische Sozialismus* ("Marxismus").

STURMTHAL, ADOLF (1943) 1951 *The Tragedy of European Labor: 1918-1939*. New York: Columbia Univ. Press.

TALMON, JACOB L. (1960) 1961 *Political Messianism: The Romantic Phase*. New York: Praeger.

TAWNEY, R. H. 1920 *The Acquisitive Society*. New York: Harcourt. → A paperback edition was published in 1948.

TAWNEY, R. H. (1931) 1965 *Equality*. 4th ed., rev. New York: Barnes & Noble.

VON MISES, LUDWIG (1922) 1959 *Socialism: An Economic and Sociological Analysis*. New ed., enl. New Haven: Yale Univ. Press. → First published as *Die Gemeinwirtschaft*.

WELLS, H. G. (1910) 1927 *The New Machiavelli*. New York: Duffield. → A paperback edition was published in 1946 by Penguin.

SOCIALIST ECONOMIC THOUGHT  
See under ECONOMIC THOUGHT.

SOCIALIZATION

- |                              |                      |
|------------------------------|----------------------|
| I. PSYCHOLOGICAL ASPECTS     | Roger V. Burton      |
| II. ANTHROPOLOGICAL ASPECTS  | John W. M. Whiting   |
| III. POLITICAL SOCIALIZATION | Fred I. Greenstein   |
| IV. ADULT SOCIALIZATION      | Orville G. Brim, Jr. |

I  
PSYCHOLOGICAL ASPECTS

This discussion of the psychological aspects of socialization will focus on the *developmental* as-

pects of individual behavior which occur in or directly involve interaction with one or more persons, i.e., in a *social* context. Any interest in the differential effects of early experiences of the infant, of child-rearing practices, of peer group influences, in sum, of any of the factors which may shape the configuration of behaviors that is considered to be personality, directs us to the subject area of social development.

Interest in this area is probably as old as almost any other human concern and is portrayed throughout written history. It is documented in Biblical accounts, the works of philosophers, autobiographies, and treatises on child-rearing practices. These accounts have been written mainly on the basis of very limited evidence, usually little more than the author's own experiences. With the development of scientific methods for studying human behavior, investigators have increasingly turned their attention to exploring all aspects of social behavior and how it develops.

**Approaches.** Methods employed to study social development are to a great extent determined by the questions the investigator asks and include almost all the techniques which have ever been used in psychological research. The approaches range from experimental to correlational, longitudinal to cross-sectional, and comprise case studies, normative and descriptive studies, small sample comparisons within a limited social group, cross-cultural analyses of many primitive societies, interviews and questionnaires, direct observations, etc. An excellent discussion of the many different approaches is provided by John Anderson (1946). Currently, an increasing interest in methods and techniques per se parallels an increasing sophistication among investigators regarding the strengths and weaknesses of these methods and the modifications required when employing them with groups differing in important respects, such as age, social background, and race. In 1960 this interest in method was manifested when the Committee on Child Development of the National Academy of Sciences-National Research Council sponsored the publication of a collection of papers presenting what was considered the wisdom of expert investigators, gained over years of using these methods (Mussen 1960). This awareness of the need for careful assessment of the appropriateness of a technique to be applied to the questions, subjects, and inferences under investigation is a healthy sign. Publication of books and papers which focus on critical and frequent foibles in the use of certain approaches and methods stimulates concern with and creative use of old and new techniques.

### Historical perspective

**Geneticism and environmentalism.** One of the earliest controversies about the different factors involved in development was concerned with the relative impact on behavior of genetic endowment and environmental influences. This controversy led to heated arguments between most early researchers of social behavior. In the work of Binet (Binet & Simon 1905) on measuring intelligence, we see how taking a position on the environment-heredity issue can become associated with the goals and purposes of research without being a necessary assumption for it. As the attempts to develop a measure of intellectual abilities which would be comparable for different ages, would be stable for individuals over time, and would predict performance in school became increasingly successful, it was difficult not to acknowledge that what was measured was something mainly genetically endowed. And the relatively successful accomplishment of this goal led many to emphasize inherited characteristics over environmental influences. This position was clearly expressed by McDougall (1908), who, despite devoting much energy to the study of social development and being one of the first to use the phrase "social psychology," incorporated into his theory a long list of instincts. This preformationist position accorded little importance to the contribution of external factors to the development of many social behaviors, such as aggression, gregariousness, and even techniques of child rearing. Gesell (1946) was another important contributor to the genetic position, with his model of reciprocal interweaving of the processes of behavioral differentiation and integration. Although some recognition of the contribution of environmental influences was made in his discussions of learning, these external factors were clearly less important than the teleological processes determining individual development. This emphasis on inherited developmental tendencies is also seen in the writings of Piaget (1923). Piaget's schemata portray definite stages, or levels, of orderly progression, each stage occurring at a particular age, supporting the position of a genetically determined sequence in the developmental process. [See DEVELOPMENTAL PSYCHOLOGY, *article on A THEORY OF DEVELOPMENT*; GENETICS, *article on GENETICS AND BEHAVIOR*; INTELLIGENCE AND INTELLIGENCE TESTING; and the biographies of BINET and GESELL.]

A similar position, but with important differences, was earlier proposed by Baldwin (1897) to account for both the mental and social develop-

ment of children. He posited that children have an inclination, apparently innate, to imitate and that imitation occurs in specific stages. This position was further developed by G. H. Mead (1934), in his description of the process whereby one learns to assume the role of another person. Although Baldwin and Mead seemed to assume an innate tendency to imitate, their presentations make external events crucial in determining much of social behavior, since the child cannot imitate a model unless it is there to copy. [See IMITATION; and the biographies of BALDWIN and MEAD.]

**Psychoanalysis.** Interest in social development was further stimulated by Freud and the psychoanalytic theorists, whose position regarding the relative importance of innate and environmental influences is a mixed one. Although the basic driving forces of behavior are assumed to be inherited, many of the manifest forms of action are acquired through associations during interpersonal situations. The importance of these external events in the development of social behavior was especially recognized by Freud, who maintained that early familial influences could determine later personality characteristics. Freud was primarily interested in dynamic, complex motivations and how these dynamically interrelated forces account for the manifest social behavior of an individual. But in his early writing (1900) he addressed himself explicitly to the question of the basic mechanism accounting for these later complex associations. The key to this system, he proposed, is the reflex arc. Basic responses of personality are conditioned by early experiences. Clearly, then, the early experiences of each individual will be paramount in shaping his social development, according to this theory.

**Behaviorism and learning.** Although it was Pavlov (1927) who investigated and established the important variables involved in the acquisition of conditioned reflexes, he did not directly study social development. It was with Watson that conditioning became all-important in accounting for social behavior, as well as for almost all other behavior. Watson (1919) proposed that a person's actions are—aside from a limited number of innate reflexes—completely the product of conditioning experiences. With this theory the pendulum had swung to the far extreme, completely away from any instinctivistic assumptions. Watson demanded that only external events and directly observable behavior be considered in building a psychological theory. With his research on children and their neurotic fears, Watson seemed very interested in demonstrating to parents the practical applications



of his theory. Although the theory's influence on the field of social development declined after a decade, the attitude toward scientific investigations of behavior through experimental methods, based on operational definitions and measurements, has been lasting. In addition, the battle between the environmentalists and nativists no longer seemed to be of such importance. Future theorists would have to acknowledge the role of environmental events, no matter what assumptions or positions they took regarding genetic contributions to behavior. [See *LEARNING, article on CLASSICAL CONDITIONING; the biographies of PAVLOV and WATSON.*]

Much of the work by learning theorists, such as Thorndike, Guthrie, Tolman, and Hull, was very relevant to social development; but it was especially at Yale University's Institute of Human Relations that direct attention was devoted to research in this area. Hull (1943) provided the beginning of a systematic, mathematical account of the principles of learning, and this general theoretical approach was applied to the study of social development by Miller and Dollard (1941) and Mowrer (1950). Their work greatly stimulated research in this area, especially in the use of experimental methods. The creative use of this theoretical orientation for explanation was demonstrated by Whiting (1941) in his interpretation of the socialization processes involved in the development of a child in a primitive culture. Studies by Sears (Sears et al. 1953) of antecedent-consequent relations in parent-child behavior have given great impetus to the study of the impact of child-rearing practices on personality development. [See *the biographies of GUTHRIE; HULL; THORNDIKE; TOLMAN.*]

**Field theory.** Kurt Lewin stimulated others to direct their attention to social development, especially where such behavior occurred in different environmental settings. Although his own work dealing with the development of social behavior was not extensive, the work of those influenced by his teaching and writing has been. (See, for example, Barker et al. 1943.) The conceptual scheme which he proposed, called field theory, emphasized the gestalt principle of considering the total situation, that is, all the factors in the environment which might affect behavior, as well as the personality characteristics of the individual. This orientation has stimulated interest in the ecological approach of studying the persisting conditions in the environment which significantly affect social development. [See *FIELD THEORY; GESTALT THEORY; and the biography of LEWIN.*]

**Operant conditioning.** One impetus for a recent surge in experimental orientation comes from the

work on operant conditioning by Skinner. Although his own research dealt mainly with animals, he has discussed the application of his approach to the development of complex human behavior (1953) and even the establishment of a utopian culture (1948). His methodological prescriptions for an experimental analysis of behavior have stimulated the application of a different approach to many areas of social development previously studied mainly by indirect routes involving circuitous interpretations. [See *LEARNING, articles on INSTRUMENTAL LEARNING and REINFORCEMENT; UTOPIANISM, article on THE DESIGN OF EXPERIMENTAL COMMUNITIES.*]

Although investigators differ in their theoretical orientations, their use of this experimental, behavioral approach, which to some degree involves a consensus about terms to use in describing research, has brought many diverse questions of social behavior under similar methodological inquiry.

### Areas of research

**Characteristics of the child.** Until the 1960s most research assessed the influence of external events on the infant as if it were a one-way process. Research in the 1960s, however, shows that neonates have different sensitivities and reactions to stimulation. Some of these differences have been related to constitutional characteristics of the infant and to the conditions of birth, such as amount of anesthesia and difficulties of delivery. Research now being conducted on infant-mother relations considers the interaction between the child and mother—both the degree to which the child influences the mother's behavior and the degree to which the mother influences the child. For example, the findings that children difficult to toilet train are insensitive to unpleasant odors and that children very sensitive to being wet are easy to night-train indicate the potential interaction between important socialization processes and the child's reactivity to relevant stimuli.

**Sex.** Studies of social development must consider the sex of the child, since the varied components of appropriate sex-role identity may be differentially influenced by many factors from the first days of life. There is evidence that some differences in early behavior are related to constitutional sex differences (Bell 1960). Among infant monkeys, for example, males actively engage in aggressive play, whereas the females tend to stay on the fringe of such activity (Harlow & Harlow 1962). In cross-cultural studies such sex differences in aggressive play also have been found to

hold for children. These sex-related behavioral proclivities are probably reinforced by society, since they conform to sex-role expectations. Parents may, even without awareness, encourage certain behaviors in their sons and other behaviors in their daughters from birth. If this speculation is justified, current and future studies of early parental treatment, where very subtle aspects of rearing practices are measured, should show differential handling for the sexes, beginning with the neonate. [See *INDIVIDUAL DIFFERENCES*, article on *SEX DIFFERENCES*.]

*Ordinal position.* Ordinal position has been related to important social behavior, such as the development of affiliative motivation. Studies (e.g., Schachter 1959) indicate that under certain kinds of stress first-borns tend to want to be with others more than do persons who were born in a later ordinal position. It is possible that the amount of attention paid to the children changes, with less attention given to individual children as the family increases. It seems likely, too, that the parents' behavior changes, because of their own differential experiences with children: as subsequent children arrive, the parents have more confidence in themselves and less anxiety about things that happen to the child. The greatest difference in treatment would most likely be between the first child and the others, but probably there are differences in handling each child. The consistency of differential treatment across families needs to be studied to determine the conditions that give rise to the effects associated with ordinal position. [See *INDIVIDUAL PSYCHOLOGY*; the biography of *ADLER*.]

*Age and maturation.* The age and maturational level of a child determine whether the occurrence of a particular event will have any salient effect on him and how permanent such an effect might be. The importance of age in relation to certain environmental conditions or learning situations will be evident throughout the following discussion.

*Early environment.* Studies of the effects of institutions on children indicate that the quality of stimulation provided by the environment may have critical consequences for the child. Research in maternal deprivation suggests that there are many important differences in the extent and quality of mother-child separation. Some immediate, measurable effects of maternal separation on the child's behavior have been demonstrated, but there is little evidence of any long-term consequences if the separation is not followed by conditions that frequently compound the event. When separation is followed by institutionalization, however, the evidence suggests that serious personality

impairment may result: deficits in motor, language, and social development. Such undesirable effects are not necessarily the result of being placed in the institution per se. Differences in the amount of stimulation given children in different institutions are associated with differential degrees of deleterious effects. Unfortunately, many of these institutions provide rather impoverished environments for infants and young children. Therefore, the immediate conditions experienced in the institution should be considered the major cause of deficits, rather than institutionalization itself. It may then be discovered that these conditions and experiences have the same effect on children never separated from their parents. The institution may tend to impose on the child a schedule which does not permit him to learn important social responses or perceive and react to caretakers and other persons in the way most children do. Researchers (e.g., Lovaas et al. 1965) are investigating ways of stimulating autistic and socially unresponsive children who have been markedly affected by such conditions to react to other persons as important stimuli in their environment. This work will attempt to modify social apathy and indiscriminate response to caretakers among these children and increase their tendency to imitate others. Work with animals that is analogous to therapeutic procedures (Schoenfeld 1950) suggests that even though the retarded or deviant behavior may be desirably modified, the person may still have a tendency to react to any future experiences of a similar nature in a dramatic fashion, responding abnormally under certain stressful conditions. [See *INFANCY*, article on *THE EFFECTS OF EARLY EXPERIENCE*.]

*Critical periods.* The question whether there is a "critical period" for institutional deprivation experiences is unsettled. There is some suggestion that the age at which institutionalization occurs has different effects: retardation is associated with early commitment, whereas children eventually diagnosed as schizophrenic tend to have been institutionalized at a later age. There is some evidence that children who have a tonsillectomy when they are between three months and three years of age are more affected than children whose tonsillectomies occur when they are either younger or older. Even within this period, the effects are lessened by preparing the child before the operation and by having a familiar adult, usually his mother, with him as much as possible during his stay in the hospital. Since some individuals who experienced severe early childhood traumata due to separation from parents are not notably disturbed in



later years, it would seem that the deprivation must be of a certain kind or quality or extend over a certain period to have any lasting effect.

*Father models and initiations.* Most of the research on separation and early deprivation focuses on the mother-child relationship and does not deal as much with the role of the father. This tendency to ignore the father's influence on the child has been true of much of the research in child development. There is support, from cross-cultural and experimental studies, for the hypothesis that a child will imitate and identify with a person who controls his important resources. This theory suggests that the relationships between parents and other persons in the child's awareness and their differential control of resources are important variables in his social development. Not only should child-rearing techniques be investigated, but research ought to consider by whom and under what conditions they are exercised.

There is evidence that a weak or absent father distorts a boy's sex identity so that he tends to be effeminate or compensatorily overmasculine. Cross-cultural studies indicate that the custom of the initiation ceremony serves to fix masculine identity in boys in societies where the father is not in the household during the first two years of the child's life (Burton & Whiting 1961). Similar conditions appear to exist in the lower-class subculture of the United States, especially in areas with delinquent gangs. Initiation rites into these gangs have characteristics similar to those of the initiation ceremonies of primitive societies, except that the rules for the boys in the primitive societies are socially desirable whereas the rules for the lower-class gang are often socially unacceptable to the society at large. Research is now needed to test whether or not any of the important elements of the initiation rites of primitive societies might function in a desirable manner for boys in our culture with the early experience of father absence or the presence of a poor father model. Furthermore, one should not look just at the undesirable side of these conditions; one also ought to consider what socially desirable behaviors—such as artistic, literary, or athletic pursuits—might result from early deprivation experiences. In the area of parental deprivation, the relationship between the absence of a male model in the home and the development of girls and the differential effects on boys and girls of maternal separation also need investigation (L. Yarrow's 1961 review contains a fuller discussion of many of these issues and studies).

*Deprivation and stimulation.* A program of comparative research stemming from the effects of

"gentling" animals to be used for experimental work shows that permanent behavioral and physiological effects are caused by different experiences during infancy. These effects are directly related to social development; animals (mainly rats) so handled tend to be less emotional, can solve certain types of problems better, are less timid, and are constitutionally more robust and socially more dominant than nongentled animals. Not as much research has been done in assessing effects of extra stimulation on children. Some early work has indicated that the later performance of children given special training for certain tasks is not very different from the performance of children who were not trained until they were more mature and could learn the tasks with a small amount of effort. Recent work does show that differential handling of infants affects their physical development: certain types of handling are associated with earlier sitting up and walking. There are also data suggesting that gentling affects human infants as well as animals, i.e., extra holding, rocking, and cuddling produces more placid and more contented babies, compared with those stimulated only for essential caretaking.

The research with animals also shows that early stimulation directly affects the growth of their endocrine system. One cross-cultural study (Landa & Whiting 1964) shows that infancy stimulation in humans has an effect on height similar to that of infancy stimulation in animals. In societies with early stressful handling (baths in water of extreme temperature, massaging, piercing of skin) the average male height is greater than in societies without such customs, even when many factors known to affect growth—such as diet, sunshine, genetics, and geographical region—are controlled. These findings should stimulate additional research into the potentially beneficial effects of certain types of infancy stress in humans. The position that any stressful stimulation during infancy should be avoided would be challenged by this research. [See STRESS.]

One sees from the research on infancy that too little stimulation is deleterious and that an extra amount of regular or "normal" stimulation promotes desirable behavior; now there is even the possibility that stressful stimulation may cause some desired consequences.

*Child-rearing practices.* Much of the research on child-rearing practices has been based on interviews and questionnaires. Most of this information has been organized along two or three conceptual dimensions: love-hostility, control-autonomy, and, possibly, calm detachment-anxious emotional in-

volverment (Becker 1964). While this is considered a useful conceptual framework for organizing research material, such two-dimensional or three-dimensional models are not meant to be any "truer" representation of the way parents interact with their children than are other models. Furthermore, these dimensions must be combined in just the right amounts, to describe what an actual parent is like. It is revealing that these data, coming from measures of verbal behavior, have a factor structure which seems very similar to the structure consistently found in studies of the underlying meaning in language: evaluation (good-bad), potency (strong-weak), and activity (fast-slow). The order or strength of the factors also is the same: good-bad or love-hostility tends to be the first factor; strong-weak or control-autonomy, the second factor; and fast-slow or high involvement-calm detachment, the third.

*Disciplinary procedures.* For general discussion, disciplinary techniques can be placed along a positive-negative continuum which appears to reflect an attitude shared by most middle-class American parents (and by research scientists). The positive techniques include rewarding, praising, and reasoning—practices considered to be psychological and love oriented and which in some ways reflect a child-centered attitude on the part of the parent. The negative methods, such as physical punishment, yelling and scolding, derogating, and threatening, reflect a more egocentric parent, in that they appear to express more of the parent's frustration over the child's behavior. Since studies in which data are obtained in interviews or questionnaires also indicate that positive disciplinary techniques are used by parents who are positive in other ways, such as providing a warm home environment and being interested in the happiness of their children and in developing their potential talents, it is difficult to winnow out just what important factors are involved in these research findings. In general, however, the positive methods of child-rearing are related to the following characteristics in the child: low aggression, low undesirable dependency, high sociability, and a high level of conscience development and of moral judgment, as reflected in a feeling of responsibility for his actions, the experiencing of guilt for deviations from prescribed standards, and the confessing of such deviations. Negative techniques are associated with the undesirable aspects of these behaviors (Becker 1964 provides a general review of this research).

These findings are frequently complicated, in that the relationships may hold only for boys or only for girls. For example, high maternal warmth

and nurturance has been associated with high conscience and leadership in boys but with low conscience and dependency in girls. Possibly, since any technique would be used in a much broader context of shaping the child in the appropriate sex role, with all the concomitant aspects of this complex concept, such as aggression, dependency, independence, differential physical and cognitive skills, demonstration of affection and other emotions, etc., the same child-rearing technique could produce quite different behavior in boys and girls. Still, it is difficult to accept, without further investigation, the idea that the same child-rearing practice by the same parent should produce opposite effects in a son and in a daughter. The evidence must be considered in light of the differential perceptions usually found for children and parents. Boys are seen as more aggressive, independent, and difficult to train than girls. Mother and father perceptions also conform to sex-role stereotypes. Mothers, as compared with fathers, are seen as more nurturant and loving and as less strict, using less physical punishment but more psychological types of discipline, and are less fear-arousing than the father. Since the same-sex parent is seen as more restrictive and frustrating than the opposite-sex parent, there are interactions in these perceptions. In addition, the father's use of physical punishment is especially frequent with boys, and the mother's use of psychological control is rated as being more appropriate for girls than for boys. These differential perceptions may account for some of the differences in the consequences of the same parent-child relationship. Many of these relationships probably reveal more about beliefs or values of the respondent (mother, father, teacher, or child) than they reflect the actual methods or personal characteristics considered important in child rearing. The contaminated research design (all data from the same respondent) used in most of these studies raises questions about the validity of the measures and the relationships between them. Expectations of investigators and types of bias within respondents have been shown to effect just such findings as those in the child-rearing research. Investigations of the degree of correspondence between measures obtained at the time child-rearing events occurred and retrospective accounts of these same events show little correlation (M. Yarrow et al. 1964). The number of significant correlations between measures of rearing techniques obtained through direct observations of parent-child interactions and measures of these same techniques obtained through interview ratings or questionnaires is no greater than chance (Zunich 1962;



Brody 1965). These data give us pause when we consider interpretations of findings in child-rearing studies. There is a trend, in this research, for investigators to cross-verify results based on indirect methods with the results of more direct measures, a trend which seems necessary to clarify some of the theoretically conflicting interpretations proposed by different researchers.

There are serious difficulties in demonstrating the relationships of child-rearing practices with later behavior, because of the many intervening events which may lessen the evidence of some direct effect. On the other hand, behavior learned by a small child may not manifest itself until a much later time, when it is appropriate. Certain questions, therefore, can be answered only through longitudinal studies, even though they are difficult and costly to execute.

**Psychoanalytic propositions.** Psychoanalytic theory stimulated interest in the effects of early socialization processes on social and personality development.

**Weaning and oral behaviors.** Freud termed the first period of development the oral stage and attributed many enduring consequences to events surrounding oral stimulation and frustration, especially the feeding process. In spite of a great deal of research dealing with types of nursing (breast or bottle, demand or scheduled) and weaning practices, there is little consistent evidence of permanent personality characteristics related to these feeding experiences. There seem to be no differential personality consequences between breast and bottle feeding; the important variable appears to be how the mother feels about it. The amount of upset associated with weaning has been shown to increase with the length of time the child nurses, up to about 12 months, at which time it begins to decrease; in other words, there is a curvilinear relationship between amount of upset and length of nursing. Cross-cultural research suggests that the amount of upset is related to adult oral anxiety, measured by the extent of attributing illness to something one has eaten. Perhaps of greater interest to parents in our society is the fact that thumbsucking has also been related to upset over weaning. Nonnutritive sucking, however, is not necessarily an index of anxiety. Research indicates that it may just as well be associated with a strongly reinforced habit as with the amount of disturbance and frustration involved in feeding and weaning. These data suggest that if only a positive experience has been associated with sucking, the response is modifiable by making some incom-

patible response more rewarding. However, thumb-sucking seems to be self-rewarding, so it will not just stop by itself—at least not for a very long time. Punishing it appears to make it more resistant to elimination. For the present, one recommendation to parents who wish to stop their child from thumbsucking is to give him a pacifier. This seems to be quickly forgotten by the child when it is too much trouble for him to find or when something interesting engages his attention. The response is then broken up in the natural course of events and disappears.

**Toilet training.** The many studies of toilet training have also tended to yield generally inconsistent and insignificant results. The traits considered characteristic of the “anal” personality, i.e., parsimony, obstinacy, orderliness, have been found to occur together in people, but the indications are that these behaviors are more likely due to direct training than to the indirect effects of toilet training. This suggests that the learning conditions established by the mother and her attitudes about her child, which may determine her toilet-training regimen, have lasting effects on social development.

**Aggression and dependency.** Two of the most thoroughly investigated areas of social development are aggression and dependency. These domains have been studied using many different methods and with both humans and animals as subjects. In general, physical punishment by parents is related to physical aggression in the child. There is evidence of limiting conditions for the overt expression of aggression, especially for girls. When fantasy and direct observations are used as measures, covert aggression is directly related to the extent of parental use of physical punishment, but overt aggression is curvilinearly related (Sears et al. 1953). Where there is no physical punishment, the child shows little aggression. As parental use of physical punishment increases, so does the child's aggressive behavior—up to a certain point. It seems that at some level of punishment overt aggressive behavior is inhibited, but covert aggression—as expressed in fantasy—is not. This information indicates that the more aggressive the parents are toward the child, the more aggression will be produced in the child's motivational system, even though overt aggression may be discouraged when the punishment is severe enough. The implication is that there might be some conditions under which this repressed hostility would become overt. Direct support for this conclusion comes from experimental studies which show that a child will imitate the aggressive behavior of a model

when the opportunity presents itself and the context indicates that no punishment for such aggression will occur (Bandura & Walters 1963). Such findings are supported by cross-cultural research which shows that societies rated high on use of painful child-rearing customs (rough handling and physical punishment) tend to have aggressive rather than benevolent gods. Other studies demonstrate that aggression can be produced in children through permissiveness toward and direct rewarding of aggressive behavior.

For dependency, the evidence again is that it may be produced through direct reinforcement. In addition, the combination of rewarding dependency when the child is young and then frustrating or punishing such behavior after it becomes part of his behavior pattern seems to produce dependency of a more rigid and persistent nature. If no punishment is associated with dependency behavior, it seems that not reinforcing it any longer will extinguish it. In light of the findings of investigations of both aggression and dependency, a reasonable hypothesis appears to be that behavior which is rewarded and encouraged in the early years and then punished when the child is older will be difficult to control and stop. Current evidence suggests that an effective procedure would be to stop rewarding the behavior the parent wants to eliminate and reward some incompatible behavior instead.

**Modifying behavior.** Research in nursery school settings has revealed that social dominance, timidity, aggression, socially desirable interactions with peers, and somewhat aberrant idiosyncratic behavior, such as crawling or walking on hands and feet rather than upright, can be modified through applying social-learning paradigms, mostly some adaptation of operant conditioning in the natural setting (Scott et al. 1967). The basic design of these experiments is to have a teacher or other adult give affectionate attention immediately following the desired behavior and not respond in any way to the undesired acts. Such studies usually demonstrate that the less-desired behavior was being maintained through the teacher's paying attention to the child (often by scolding) only when he was doing what she did not want him to do. By contrast, the teacher tends to ignore him when he behaves well. Rarely have these studies investigated the combination of positive and negative reinforcement for desired and undesired behavior, respectively. Since there is evidence from learning experiments that when the two responses are mutually exclusive such differential rewarding and punishing is the most efficient teaching paradigm,

it will be important to know how such contingencies work in a natural, everyday teaching situation. The hypothesis stated above for aggression and dependency implies that once behavior has become part of the child's repertory of responding, the introduction of punishment, while it may inhibit the behavior temporarily, will actually make it more resistant to modification. However, if there is a mutually exclusive type of behavior readily available for the parents to reward, a combination of disapproval of the undesired activity and approval of the acceptable behavior may be more effective than trying to modify the behavior through approval alone. [See *LEARNING, especially the articles on INSTRUMENTAL LEARNING and REINFORCEMENT.*]

**Labeling.** There is experimental evidence that clear labeling of behavior helps a child discriminate in one situation and then transfer his learning to a new situation. This finding suggests that parents and teachers will be more successful in developing certain desired social behavior in their children by clearly labeling the behavior they are rewarding and punishing. This interpretation implies that the more they provide the child with cognitive support for their approval and disapproval, the more the child will understand how he should act in a quite different place but where the abstract aspect of the situation is the same.

**Consistency.** The desirability of consistency in child rearing is often stated. However, there are many kinds of consistency possible, ranging from the evaluation of certain behavior as always good or always bad to the unvarying use of certain techniques for rewarding or punishing. Most research on consistency has addressed itself to consistency of the former type. Inconsistent or capricious discipline is associated with socially undesirable behavior, as is apparent from studies of delinquents and their parents. There is evidence which suggests that some ways of being inconsistent would produce desired results. First, if the parents always use the same kind of reward (or punishment), the child might become satiated with it (or adapt to it) and its power to reinforce would be attenuated. This suggests that to maintain effectiveness, rewards and punishments for the same type of behavior should be varied. Studies also show that once behavior is well established, making rewards or punishments unpredictable for the child will maintain the behavior in the absence of the parent better than if the parent is consistent in rewarding every time the child acts well or in punishing every time he is bad. The important distinction here is



that a type of behavior, when it is reinforced, is either always rewarded or always punished, in contrast to the undesirable kind of inconsistency, where the same behavior is sometimes rewarded and sometimes punished.

**Peer-group influences.** The first peer group for many children consists of their siblings. The sex of these siblings is important to the child's social development since older sibs may have much influence over him, both directly, through rewarding and punishing his behavior, and more indirectly, through providing a model. These experiences would be expected to affect his later social relations (Koch 1960). There is evidence that a later preference for same-sex or opposite-sex playmates is related to whether the child has had opposite-sex sibs or not, relatively greater preference for opposite-sex playmates being associated with having opposite-sex sibs (Koch 1957).

Studies have assessed the effects of nursery school and kindergarten attendance on social development, based on the assumption that such effects stem from the child's association with other children his age. The results are inconsistent: some show positive associations with desirable social development, and others show no relationship or even show some indication, in the investigator's opinion, that the child in nursery school may be habituated in his behavior too early. To a great extent the variability of results seems due to differences in the kinds of measures and behaviors employed in these studies. At this time there seems to be little that can clearly be stated about empirical results or that would lead us to even a tentative conclusion about the differential effect of preschool attendance *per se* on a child's social development. Differences between various nursery school settings do have differential effects on the child's social behavior, but investigations to determine the important variables distinguishing the settings still need to be made. Some studies clearly show that social behavior of individual children, as well as group differences, are modified by certain training regimes and specific procedures. Ascendancy and submission, social participation, leadership, and positive and negative social interaction have been the behaviors modified in the direction considered desirable.

Peer influences on social behavior have been shown to affect aggression, honesty, expressed values, leadership, participation in diverse extra-curricular activities, friendship choices, and identity with a group and its activities. These differences in behavior have been related to whether the

child was a member of a particular clique, club, classroom, school, or social class. For some of these associations it is difficult to determine how much to attribute to peer influences and how much to more basic factors, such as family influences, which would determine the child's assignment to these different social units. However, it seems clear that the peer group does influence the social behavior of the child, even if such influence is only to reinforce behavior which was already a part of the child's repertory. Interesting demonstrations of peer influences on social development come from studies of the *kibbutzim* and of Soviet methods of character training. Analyses of these peer settings indicate that the peer group does act as a powerful socializing agent in rewarding and censuring and successfully shaping behavior of its members in conformity with the group's standards. The evidence available for the Soviet methods of education comes mainly from analyses of their written works, with some support coming from interviews with teachers and a limited amount of direct observation in the schools (Bronfenbrenner 1962). More extended research of socialization has been possible in the *kibbutzim* of Israel. Direct observations in the nursery indicate that the peer group is important in developing socially desirable habits of eating, toileting, sex, independence, and aggression (Faigin 1958).

The sex and age of the child are important attributes in determining what peer grouping he becomes a member of and how the peer group acts toward him. Cross-cultural studies indicate that from the earliest years the sexes are treated differently by their parents. Other studies indicate that the teaching of behaviors appropriate to the different sex roles, which is begun in the home, is reinforced by the peer groups. Sex seems to be the major determinant in early peer grouping, where the members engage with a vengeance in sex-typed play and eschew behavior considered to be characteristic of the opposite sex. With increasing age, individual friendships become stronger, the role of the individual in the group becomes more distinctive, organized social activity increases, and the individual's identity with the group and the degree to which he is controlled by the group's sanctions increases. Even for groups formed for just a short time, there is evidence that reputations and expectations of an individual's behavior are quickly formed and are quite stable. Such social perceptions by the peer group appear to be strong determinants in shaping the member's behavior to conform to his reputation and the expectations the

others have for him (Campbell's 1964 review provides fuller coverage of this topic).

**Other influences.** Since the average American child spends much time watching television and motion pictures, the importance of these media for teaching values and different kinds of behavior is obvious. A program of research in social learning by Bandura and Walters (1963) shows that children will imitate an aggressive model in a motion picture. Watching aggression in a movie also elicits preference for and participation in aggressive play. For adults there is some evidence that, when an individual is already aggressively aroused, watching an aggressive movie acts as a catharsis. For children, however, far from acting as a fantasy reduction of aggressive impulses, the viewing acts to arouse aggression in them. These studies indicate that aggression is more easily aroused and elicited in children than are other forms of behavior. There is some indication that adolescents who watch television programs depicting many stresses and strains of adult life are likely to be more anxious and dependent on their families than are children who do not watch such programs. Television viewing has been associated with a tendency to stereotype nationalities and at the same time with a tendency not to ascribe value judgments to minorities. The evidence suggests that the mass media may have both very desirable and undesirable influences on social development. The effects of viewing appear to be found mainly when the behavior learned is reinforced in the home or is not dissonant with the home values. Still, the implication is that parents would do well to exert some supervision over the viewing choices of their children.

**Reading material.** Little evidence of the effects of reading material on behavior is available. Studies have shown that the reading of an inordinate number of comic books, especially those with highly aggressive content, may be a symptom of emotional problems in a child and is more characteristic of delinquent children than nondelinquents. But the causal relationship is not clear, since the disturbed child may turn to the fantasy material because of his anxieties. Content analyses of recommended children's literature indicate rather consistent portrayal of certain roles having certain behavioral traits. This probably reinforces what the child is learning in the home, school, and peer groups about these roles. What the child reads may, nevertheless, teach him certain attitudes, values, and behaviors, which may later be manifested as overt behavior in appropriate situations. But at this time

there seems to be no clear evidence in support of this hypothesis.

### Trends

Several main currents in the flow of research in social development may be noted. One of these, which has a well-established tradition and has always been an integral part of this area of investigation, is the application of learning principles from the experimental laboratory to the more complex areas of human behavior. This paradigm has been especially prominent in recent research investigating some of the important parameters involved in the development of social behavior. The way of thinking about stimuli, responses, reinforcements, and the many other factors found important in experimental studies has been employed both in the design and in the conceptualization of studies of social behavior in nonexperimental settings, and this trend is increasing. This stamp has been felt in no small way in the tightening up of definitions of the behavior and other variables under study. The influence seems to be most apparent in the following kinds of research: (1) observational studies, where data are obtained in a sequential manner in order to ascertain the potential effects of immediate events on behavior; (2) applied uses in therapy and in the modification of behavior of institutionalized persons; and (3) the creative use of experimental designs in natural settings, where the changes in behavior produced by the experimenter's control of the environment are investigated.

Another dominant trend is in the self-conscious awareness of methods and measuring techniques. There are signs of uneasiness about the labels that have been applied to certain measures which have not been carefully tied to actual behavior. There is increasing use of research designs to assess the correspondence between indirect and direct measures of attitudes and beliefs and their overt behavioral manifestations. The indications are that information from all these levels will provide a fuller understanding of the processes involved in behavior and its development. New techniques are being devised to record and analyze sequential events, in order to assess the effects of different environmental contingencies in determining actions. Obtaining such data in turn demands the development of new kinds of multivariate analytical models, especially for analyzing interaction sequences in a setting such as the home or school.

Another important trend stems from studies of the newborn. These studies have forced researchers



to consider possible genetic factors involved in social development. Emphasis is placed on studying the mother-infant relationship as a truly dyadic one, where the infant is shaping the mother's behavior as well as being affected by her. The concept of "critical periods" has been given special consideration in this area of research. The question of whether an experience has differential effects depending on the maturational age of the child is now being extended from studies of the infancy period to studies involving later ages. [See IMPRINTING.]

ROGER V. BURTON

[Directly related are the entries ADOLESCENCE; DEVELOPMENTAL PSYCHOLOGY; INFANCY; PERSONALITY, article on PERSONALITY DEVELOPMENT. Other relevant material may be found in AGGRESSION; IMITATION; LEARNING, especially the article on LEARNING IN CHILDREN; PSYCHOANALYSIS.]

#### BIBLIOGRAPHY

- ANDERSON, JOHN (1946) 1954 *Methods of Child Psychology*. Pages 1-59 in Leonard Carmichael (editor), *Manual of Child Psychology*. 2d ed. New York: Wiley.
- BALDWIN, JAMES M. (1897) 1906 *Social and Ethical Interpretations in Mental Development: A Study in Social Psychology*. 4th ed., rev. & enl. New York: Macmillan.
- BANDURA, ALBERT; and WALTERS, R. H. 1963 *Social Learning and Personality Development*. New York: Holt.
- BARKER, ROGER G.; KOUNIN, J.; and WRIGHT, H. F. (editors) 1943 *Child Behavior and Development*. New York: McGraw-Hill.
- BECKER, WESLEY C. 1964 *Consequences of Different Kinds of Parental Discipline*. Volume 1, pages 169-208 in Martin L. Hoffman and Lois W. Hoffman (editors), *Review of Child Development Research*. New York: Russell Sage Foundation.
- BELL, RICHARD Q. 1960 *Relations Between Behavior Manifestations in the Human Neonate*. *Child Development* 31:463-477.
- BINET, ALFRED; and SIMON, TH. 1905 *Méthodes nouvelles pour le diagnostic du niveau intellectuel des anormaux*. *Année psychologique* 11:191-244.
- BRODY, GRACE F. 1965 *Relationship Between Maternal Attitudes and Behavior*. *Journal of Personality and Social Psychology* 2:317-323.
- BRONFENBRENNER, URIZ 1962 *Soviet Methods of Character Education: Some Implications for Research*. *American Psychologist* 17:550-564.
- BURTON, ROGER V.; and WHITING, J. W. M. 1961 *The Absent Father and Cross-sex Identity*. *Merrill-Palmer Quarterly of Behavior and Development* 7:85-95.
- CAMPBELL, JOHN D. 1964 *Peer Relations in Childhood*. Volume 1, pages 289-322 in Martin L. Hoffman and Lois W. Hoffman (editors), *Review of Child Development Research*. New York: Russell Sage Foundation.
- FAIGIN, HELEN 1958 *Social Behavior of Young Children in the Kibbutz*. *Journal of Abnormal and Social Psychology* 56:117-129.
- FREUD, SIGMUND (1900) 1953 *The Standard Edition of the Complete Psychological Works of Sigmund Freud*. Volumes 4 and 5: *The Interpretation of Dreams*. London: Hogarth; New York: Macmillan.
- GESELL, ARNOLD (1946) 1954 *The Ontogenesis of Infant Behavior*. Pages 335-373 in Leonard Carmichael (editor), *Manual of Child Psychology*. 2d ed. New York: Wiley.
- HARLOW, HARRY F.; and HARLOW, MARGARET K. 1962 *Social Deprivation in Monkeys*. *Scientific American* 207, Nov.:136-146.
- HULL, CLARK L. 1943 *Principles of Behavior: An Introduction to Behavior Theory*. New York: Appleton.
- KOCH, HELEN L. 1957 *The Relation in Young Children Between Characteristics of Their Playmates and Certain Attributes of Their Siblings*. *Child Development* 28:175-202.
- KOCH, HELEN L. 1960 *The Relation of Certain Formal Attributes of Siblings to Attitudes Held Toward Each Other and Toward Their Parents*. *Society for Research in Child Development, Monographs* 25, no. 4.
- LANDAUER, THOMAS K.; and WHITING, J. W. M. 1964 *Infantile Stimulation and Adult Stature of Human Males*. *American Anthropologist New Series* 66:1007-1028.
- LOVAAS, O. IVAR et al. 1965 *Experimental Studies in Childhood Schizophrenia: Analysis of Self-destructive Behavior*. *Journal of Experimental Child Psychology* 2:67-84.
- MCDUGALL, WILLIAM (1908) 1950 *An Introduction to Social Psychology*. 30th ed. London: Methuen. → A paperback edition was published in 1960 by Barnes and Noble.
- MEAD, GEORGE H. (1934) 1963 *Mind, Self and Society From the Standpoint of a Social Behaviorist*. Edited by Charles W. Morris. Univ. of Chicago Press. → Published posthumously.
- MILLER, NEAL E.; and DOLLARD, JOHN 1941 *Social Learning and Imitation*. New Haven: Yale Univ. Press; Oxford Univ. Press.
- MOVWER, O. HOBART 1950 *Learning Theory and Personality Dynamics*. New York: Ronald.
- MUSSEN, PAUL H. (editor) 1960 *Handbook of Research Methods in Child Development*. New York: Wiley.
- PAVLOV, IVAN P. (1927) 1960 *Conditioned Reflexes: An Investigation of the Physiological Activity of the Cerebral Cortex*. New York: Dover. → First published as *Lektsii o rabote bol'shikh polusharii golovnogo mozga*.
- PIAGET, JEAN (1923) 1959 *The Language and Thought of the Child*. 3d ed., rev. New York: Humanities Press. → First published as *Le langage et la pensée chez l'enfant*.
- SCHACHTER, STANLEY 1959 *The Psychology of Affiliation: Experimental Studies of the Sources of Gregariousness*. *Stanford Studies in Psychology*, No. 1. Stanford Univ. Press.
- SCHOENFELD, WILLIAM N. 1950 *An Experimental Approach to Anxiety, Escape and Avoidance Behavior*. Pages 70-99 in Paul H. Hoch and Joseph Zubin (editors), *Anxiety*. New York: Grune.
- SCOTT, PHYLLIS M.; BURTON, R. V.; and YARROW, M. R. 1967 *Social Reinforcement Under Natural Conditions*. *Child Development* 38, no. 1.
- SEARS, R. R. et al. 1953 *Some Child Rearing Antecedents of Aggression and Dependency in Young Children*. *Genetic Psychology Monographs* 47:135-236.
- SKINNER, B. F. 1948 *Walden Two*. New York: Macmillan.
- SKINNER, B. F. 1953 *Science and Human Behavior*. New York: Macmillan.

- WATSON, JOHN B. (1919) 1929 *Psychology From the Standpoint of a Behaviorist*. 3d ed., rev. Philadelphia and London: Lippincott.
- WHITING, J. W. M. (1941) 1942 *Becoming a Kwoma: Teaching and Learning in a New Guinea Tribe*. New Haven: Yale Univ. Press.
- YARROW, LEON J. 1961 Maternal Deprivation: Toward an Empirical and Conceptual Reevaluation. *Psychological Bulletin* 58:459-490.
- YARROW, MARIAN R.; CAMPBELL, J. D.; and BURTON, R. V. 1964 Reliability of Maternal Retrospection: A Preliminary Report. *Family Process* 3:207-218.
- ZUNICH, MICHAEL 1962 Relationship Between Maternal Behavior and Attitudes Toward Children. *Journal of Genetic Psychology* 100:155-165.

## II

## ANTHROPOLOGICAL ASPECTS

"Socialization" gained currency in the 1930s as a term denoting the process by which culture is transmitted from one generation to the next. Dollard described the process as "an account of how a new person is added to the group and becomes an adult capable of meeting the traditional expectations of his society for a person of his sex and age" (1935, chapter 1). Although a review article by Irvin Child entitled "Socialization" (1954) signaled its formal acceptance, persons doing research on this process have never been too happy with the term—in part because of its ambiguous connotations and in part because it suggests that the concept is limited to the learning of social roles. This implied exclusion of the transmission of beliefs, values, and other cognitive aspects of culture led Kluckhohn (1939) to suggest "culturalization" and Herskovits (1948) to propose "enculturation" as alternative terms. Although these terms, especially the latter, have much to recommend them, they have not been widely used. Nor has "socialization" been universally accepted. "Child rearing" seems to be gaining some favor as an alternative (Sears et al. 1957; B. Whiting 1963), but it also fails to suggest cultural transmission.

## Psychoanalytic and cultural approaches

The writings of Freud turned the attention of anthropologists to the study of socialization. Although culture is by definition (Tylor 1871) learned anew by each generation, the study of this process was almost entirely neglected by anthropologists until after 1925. Up to this time, American anthropologists were primarily concerned with the history of cultures and with describing how traits were diffused or borrowed. For example, in their ethnographies a description of a cradle board detailed sufficiently so that it could be identified as either a "Plains type" or a "Plateau type" was often

the only material bearing in any way on infancy or childhood. In Europe and England, where evolutionary theories were in vogue, there was even less concern with this topic.

The early writings of Freud were little noticed by anthropologists, but when *Totem and Taboo* appeared in 1913, this invasion of the ethnological domain could not be ignored by them. For the most part, the initial response of anthropologists was negative. Kroeber (1920) wrote a critical review of the work in the *American Anthropologist*, and Malinowski (1927), in an attack upon Freudian theory, argued that the Oedipus complex did not occur in the Trobriand Islands. It was Edward Sapir who anticipated the view of psychoanalytic theory that was to be taken by many anthropologists:

Those who are profoundly convinced of the epoch-making importance of the psychological mechanisms revealed by Freud and, even more, of the extraordinary suggestiveness of numerous lines of inquiry opened up by psychoanalysis, without, at the same time, being blind to criticisms that need to be made of certain psychoanalytic theory, can only hope and pray that this not altogether healthy overpopularity of the subject prove no hindrance to the study of the perplexing problems with which the Freudian psychology bristles. ([1917] 1949, p. 522)

Margaret Mead, setting sail for Samoa in 1925, was the first anthropologist to engage in field work with the avowed intention of studying an aspect of socialization. In this first study, she was not strongly influenced by psychoanalytic theory. In her report on the life of adolescent girls (1928) she did not talk of Oedipal complexes or oral fixations but rather about the everyday life of these girls, paying particular attention to the ways in which their lives contrasted with those of American girls of the same age. Her description of growing up in New Guinea (1930) was also more anthropological than psychoanalytic in its conception, but her later study of three contrasting cultures in New Guinea (1935) clearly showed Freudian influence.

During this early period two psychoanalysts, Géza Róheim and Erik Erikson, did field work in non-Western societies; the former worked mainly in Australia and Melanesia (Normanby Island), and the latter studied the American Indians (Sioux and Yurok). Both of them reported child-rearing practices and interpreted some of the rituals, values, and customs as the consequence of unconscious conflicts engendered by these practices. Róheim was so arbitrary and dogmatic in his interpretations that he antagonized many anthropologists and gave the study of socialization a bad



name among the more conservative members of the profession. Erikson's interpretations were, by contrast, sensitive, illuminating, and nondogmatic—a fact which did much to counteract the influence of Róheim (see Erikson 1950).

In 1932–1933 Edward Sapir and John Dollard gave a joint seminar on culture and personality at Yale University. Dollard, a sociologist, had just returned from Germany, where he had received psychoanalytic training. The combination of the cultural and psychoanalytic approaches for the study of personality development was clearly expressed by Dollard soon after this in his *Criteria for the Life History* (1935). Partly as a consequence of this seminar, a number of life histories of preliterate men and women were published (see Dyk 1938; Ford 1941; Talayesva 1942).

A few years later Abram Kardiner, a psychoanalyst, and Ralph Linton, an anthropologist, instituted another seminar on personality development, held this time at Columbia University, which resulted in the publication of two important volumes: *The Individual and His Society* (Kardiner 1939) and *The Psychological Frontiers of Society* (Kardiner 1945). These consisted of case studies of a number of societies whose projective systems were interpreted in the light of their "basic disciplines," for example, weaning, toilet training, and the control of sex, dependence, and aggression. The cultural material was supplied by an anthropologist with a firsthand knowledge of the case; in most instances the field work had been done prior to the seminar. Cora DuBois' work on the Alorese, reported both in the second volume and in a separate monograph (1944), was a landmark in the study of socialization from the combined viewpoints of psychoanalysis and cultural anthropology.

The influence of psychoanalytic thinking upon anthropologists that occurred mainly in the decade 1925–1935 resulted in a dramatic change in ethnographic research. Most general ethnographies written after 1930 included descriptions of infancy and childhood. Furthermore, field studies explicitly focusing on the socialization process were undertaken. Child-rearing practices were even used to interpret the "national character" of modern nations [see NATIONAL CHARACTER].

### Learning theory and socialization studies

In the mid-1930s, learning theory was added as a third component in the study of socialization. This came about as a consequence of a seminar at the Institute of Human Relations at Yale University. Among the participants in this group were John Dollard and Earl Zinn, representing the psy-

choanalytic viewpoint; C. S. Ford, B. Malinowski, G. P. Murdock, and John W. M. Whiting, representing anthropology; and Leonard Doob, Carl Hovland, Clark Hull, O. Hobart Mowrer, Neal Miller, and Robert Sears, representing learning and behavior theory.

Since both the psychoanalytic and the cultural approaches imply learning, the early studies of socialization had an implicit and usually common-sense theory of learning which seemed to serve the purpose well enough at the ethnographic level. Thus Mead described an aspect of Samoan education as follows:

From birth until the age of four or five a child's education is exceedingly simple. They must be housebroken, a matter made more difficult by an habitual indifference to the activities of very small children. They must learn to sit or crawl within the house and never to stand upright unless it is absolutely necessary; never to address an adult in a standing position; to stay out of the sun; not to tangle the strands of the weaver; not to scatter the cut-up cocoanut which is spread out to dry; to keep their scant loin cloths at least nominally fastened to their persons; to treat fire and knives with proper caution; not to touch the kava bowl, or the kava cup, and, if their father is a chief, not to crawl on his bed-place when he is by. These are really simply a series of avoidances, enforced by occasional cuffs and a deal of exasperated shouting and ineffectual conversation. (1928, pp. 22–23)

The attempt by John W. M. Whiting (1941) to apply a more formal version of learning theory to the socialization process of the Kwoma of New Guinea was, as Gladwin comments, "so literal that it was almost a tour de force. The exercise has therefore not been repeated" (1961, p. 153). Certainly the Kwoma data were in no sense the test of a theory, but they did provide a vehicle for presenting a view of certain aspects of the socialization process. In this study the role of the socializing agent was stressed, as indicated by the following quotation:

A Kwoma child learns but a small part of his cultural habits by free trial and error, that is, without some member of his society guiding and directing him. Were he to do so, he would learn those habits which were most rewarding to him and to him alone. This, however, is not what actually happens. He is forced to learn, not the habits which might be most rewarding to him alone, but the habits which are specified in the culture as being best. . . . Thus an essential set of conditions for social learning is the behavior of the socializing agents. . . .

Before the analysis of teaching techniques is undertaken, several terms which are to be used should be defined. The terms "teacher" and "pupil" will be employed in a special sense, the former to mean anyone who attempts to change the habit structure of another

person, the latter, any person whose habit structure is being so changed. Although Kwoma parents most frequently play the role of teacher, with their children as pupils, the teacher-pupil relation is not by any means restricted to them. Co-mothers, siblings, paternal uncles, and aunts, and other relatives frequently play the role of teacher as well. . . .

Unless the essentials for learning are already supplied by the interaction of the environment and the pupil, they must be supplied by the teacher. In other words, when such conditions are not already present, the teacher, in order to change the habit structure of the pupil in the desired manner, must provide motivation, guidance, and reward. Each of the various teaching techniques employed at Kwoma may be classed under one or another of these three categories. The following table presents this classification:

1. Providing motivation
  - a. punishing
  - b. scolding
  - c. threatening
  - d. warning
  - e. inciting
2. Providing guidance
  - a. leading
  - b. instructing
  - c. demonstrating
3. Providing reward
  - a. giving
  - b. helping
  - c. praising

(J. Whiting 1941, pp. 177-178; 180)

An analysis of the method by which supernatural beliefs were transmitted in Kwoma society was also presented. This analysis was intended as an antidote both to some of the more exaggerated Freudian interpretations, which seemed to imply that the so-called projective systems of a culture were re-invented each generation from the neurotic fantasies of its members, and to the bizarre "racial unconscious" hypothesis, which held that religious beliefs were genetically transmitted.

Since learning theory was not especially useful as an interpretive device, the Kwoma example was not followed. However, as one of the bases for developing the so-called complex mechanisms of social learning, it was invaluable. Another and more profound influence of learning theory was methodological rather than substantive.

Learning and behavior theory was developed in the laboratory rather than on the "couch" or in the "field," and therefore the method used was experimental rather than descriptive and interpretive. The stating of hypotheses so that they can be tested or jeopardized is crucial to a laboratory science. Although the procedures of the controlled experiment are inappropriate for the study of socialization, correlational studies both within and across

cultures are quite feasible. The effect of work by the experimental-learning theorists was to introduce hypothesis testing, case counting, and the use of statistics to the study of socialization. John W. M. Whiting and Irvin Child expressed this development in their cross-cultural study of socialization:

Most previous work in culture and personality has been oriented primarily toward seeking concrete understanding of specific cases. Our work in contrast is oriented toward testing generalized hypotheses applicable to any case. The difference is not between interest and lack of interest in general hypotheses. The student who is primarily oriented toward full interpretation of the individual case inevitably makes use of general hypotheses as an interpretive device. But he is inclined to take their validity for granted and use them simply as tools which contribute to the understanding of the concrete case. We are not willing to take the validity of any hypothesis for granted until we see adequate evidence to support it, and we are willing to leave to the future the task of applying validated hypotheses to the interpretation of specific cases. (1953, p. 5)

### The comparative study of socialization

**Socialization as cause and as effect.** The Whiting and Child study used data from 75 mostly preliterate societies for which sufficient descriptive data on the socialization process were available. The hypotheses tested were derived from Freudian theory but modified by learning theory and adapted to the categories of cultural anthropology in such a way that they were testable.

In this study Whiting and Child made a distinction between the maintenance systems and the projective systems of a culture. The former include "the economic, political, and social organization of a society—the basic customs surrounding nourishment, sheltering, and the protection of its members" (Whiting & Child 1953, p. 310); the projection systems (Kardiner 1945) include magic, art, and religion. In any culture, socialization was assumed to be an *effect* of the maintenance systems and a *cause* of the projective systems. Thus, this model assumes that insofar as socialization affects the habit structure or personality of the members of a society, it becomes one of the mechanisms by which culture is integrated.

Following this study by Whiting and Child, a fairly large number of cross-cultural studies were undertaken to test hypotheses concerning either the effect of the maintenance systems upon socialization or the effect of socialization upon magical beliefs and rituals (these are reviewed in J. Whiting 1961).



Findings relating socialization to the projective systems include the following: the treatment of infants in a society is predictive of beliefs about the nature of the gods (Spiro & D'Andrade 1958; Lambert et al. 1959; J. Whiting 1959a); the severity of socialization leads to complexity in decorative art (Barry 1957); achievement imagery in folk tales is predictable from the age and severity of independence training (McClelland & Friedman 1952; however, in a replication of this study by Child et al. 1958, this relationship was not supported by the data); last, that an exclusive relationship between mother and son during infancy is associated with male initiation rites at puberty (J. Whiting et al. 1958; Burton & Whiting 1961).

Studies purporting to show that socialization is an effect of maintenance-system variables have found the following associations: the amount of indulgence during infancy is predictable from the number of adult females in the household—large extended family households being the most indulgent and households consisting of a mother and child alone being the least indulgent (J. Whiting 1961); extended family households are also the most severe in controlling the aggressions of their children (J. Whiting 1959b); the type of subsistence economy is predictive of the strength of socialization pressure to train the child toward being compliant rather than assertive, the strongest pressure being found among pastoral societies and the weakest among the hunters and gatherers (Barry et al. 1959).

**Comparative case studies.** Paralleling the type of cross-cultural studies reported above was a series of comparative case studies using the so-called method of controlled comparison. Margaret Mead's investigation of three societies in New Guinea—the Arapesh, the Mundugumor, and the Tchambuli—was the first of these (1935). In this study she concluded that the personalities of the two sexes are socially produced and may vary to the extent that the biological stereotype may actually be reversed.

A few years later the first large interdisciplinary project was undertaken. This was the Indian Education Research Project jointly undertaken by the Committee on Human Development of the University of Chicago and the United States Bureau of Indian Affairs. A long list of notables from the fields of anthropology, psychology, and education constituted the advisory committee. This ambitious project undertook the study of socialization and its effects upon personality in each of five societies: the Hopi, Navajo, Papago, Sioux, and Zuni. Standard tests such as the Rorschach, the Thematic Apperception Test, Stewart's Emotional Response

Test, Goodenough's Draw-a-Man Test, the Arthur Point Scale of Performance, and Bavelas' Moral Ideology Test were given to children, and the societies were compared on the basis of the test results. The major results of the project, however, were a series of descriptive monographs (Leighton & Kluckhohn 1947; Thompson & Joseph 1944; Macgregor 1946; Joseph et al. 1949). Although each of these was written by a different person, they provided far more comparable data than those which are usually available for the large sample cross-cultural studies.

In the above studies the society is the basic unit of comparison. Although in the Indian Education Project tests were given to individual children, the only intra-cultural comparison made was between subgroups judged to be more or less traditional.

Another project in which several societies were simultaneously studied—the Coordinated Investigation of Micronesian Anthropology—was undertaken shortly after the end of World War II. Although this project was not specifically addressed to studies of child rearing or personality development, several excellent studies on these topics resulted, notably one on Ifaluk (Burrows & Spiro 1953) and one on Truk (Gladwin & Sarason 1953).

The advantage of these projects, as compared with the cross-cultural studies based on large samples, was that they provided case studies in greater depth and with greater comparability; however, they suffered from providing too few cases to enable a confident interpretation of observed differences.

**Replication studies.** The next step in research-design efficiency was taken by Prothro (1961), who went further than using standard instruments for comparative purposes. He set himself the task of replicating, in Lebanon, a study of child-rearing practices that had been carried out in a suburb of Boston by Sears, Maccoby, and Levin (1957). The same interview was used as a basic research instrument. Some of Prothro's findings confirmed the Boston study, and some did not:

If we compare the children who did have feeding difficulty with those who did not, we find that the first conclusion of the American study is confirmed by our data. Feeding problems for Arab and Armenian children were not linked to infant weaning problems or to age at weaning. The feeding problems of the five-year-old—New Englander or Middle Easterner—are not linked to problems of feeding and weaning in infancy.

The relationship of maternal warmth to feeding problems is, however, the opposite of that found in America. The warmer mothers were more likely to report feeding problems. Of those mothers whose children had feeding problems, 47 per cent were high in warmth and 37 per cent were low in warmth. Of mothers not

reporting any feeding problems, 35 per cent were high in warmth and 45 per cent were low in warmth. This distribution . . . could not have occurred only by chance. Maternal warmth was associated with more feeding problems. . . .

It is not difficult to adduce post hoc explanations for the difference between our findings and those in America. American mothers generally try to see that their children eat enough, and eat a "proper" diet. Warm and cold mothers alike would concern themselves with what the child ate, and the warmth or coldness would express itself in the way in which they tried to achieve this goal. In Lebanon, however, there is less attention given to the diet of the toddler, so that greater warmth might express itself in greater concern for diet, more efforts to insure the eating of "healthful" foods, and more awareness of any balky or "finicky" habits that did exist. Thus greater warmth might produce pressures comparable to those found in America. Whether this explanation is correct or not, we can be certain of one fact—the [negative] correlation found in the United States between maternal warmth and feeding problems does not obtain in Lebanon. Warmer mothers reported more such problems. (1961, pp. 83–84)

As can be seen from the above quotation, the replication strategy is powerful when the results of various studies confirm one another, but when they do not, further replications are required. Large-scale comparative studies with carefully controlled methods of data collection and replication and a large enough sample of societies to adequately jeopardize a hypothesis have yet to be undertaken.

**Subsystem replication.** An alternative strategy that Roberts and Sutton-Smith have called subsystem replication has much merit. This design requires that any hypothesis to be tested be put in double jeopardy, that is, tested on the basis of cultural differences among an adequate sample of societies and also on either subgroup or individual differences within one or more of these societies. Studies involving socialization variables and using this strategy have been done by John W. M. Whiting (1954; Whiting et al. 1966a) and Roberts and Sutton-Smith (1962).

The most elaborate attempt so far to use the subsystem-replication strategy was the so-called Six Culture Project, which was directed by Irvin Child, William Lambert, and John W. M. Whiting. For this study teams of investigators underwent training in standard methods of interviewing and observing children for a year before going into the field. At the end of this training period the senior investigators and the field teams prepared a document, *Field Guide for the Study of Socialization* (Whiting et al. 1966b), for the purpose of maximizing the uniformity of their data collection. In

addition, Beatrice Whiting acted as a clearinghouse during the period of field work and attempted to solve unanticipated field-work problems in a way that would maximize uniformity.

The design of this study called for collection of cultural data on each of the six societies, using standard ethnographic techniques for comparison at the societal level. For subsystem replication, a sample of 24 children from the ages of three to ten, balanced by sex and age, was chosen at random from the census of all children in the community selected for study. These children together with their parents were chosen for intensive study. (See B. Whiting 1963 for a description of the cultures and the child-rearing practices and Minturn & Lambert 1964 for a factor analysis of standard interviews with each of the sample mothers.)

Although the effort to insure comparability of data collection represented by this study is a step forward in the comparative study of socialization, there is still a long way to go before satisfactory standards are achieved. Perhaps the next step will be the collection of a limited amount of data relevant to the testing of a specific hypothesis, carried out by a single person or team, on an appropriate sample of societies.

Anthropologists, stimulated by the work of Freud, have in increasing numbers turned their attention to the problem of cultural transmission. Ethnographies detailing this process, almost wholly absent in 1920, are now a commonplace. This has added what might be termed a "third dimension" to the description of a culture. More importantly, it has introduced an appreciation that the "natural man" assumption is inadequate for a complete understanding of cultural integration. The evidence from comparative studies of socialization indicates that man as a bearer of culture is not just a primate who thinks logically or prelogically and whose feelings are based upon primary needs such as hunger, thirst, sex, fear, and the like; in addition, as a consequence of the child-rearing methods of his society, man has blind spots and distortions in his cognitive processes and specific anxieties which may form the emotional underpinning and *raison d'être* for elaborate rituals and magico-religious beliefs.

JOHN W. M. WHITING

[See also CULTURE AND PERSONALITY.]

#### BIBLIOGRAPHY

- BARRY, HERBERT 1957 Relationships Between Child Training and the Pictorial Arts. *Journal of Abnormal and Social Psychology* 54:380–383.



- BARRY, HERBERT; CHILD, IRVIN L.; and BACON, MARGARET K. 1959 Relation of Child Training to Subsistence Economy. *American Anthropologist* New Series 61:51-63.
- BURROWS, EDWIN G.; and SPIRO, MELFORD (1953) 1957 *An Atoll Culture: Ethnography of Ifaluk in the Central Carolines*. 2d ed. New Haven: Human Relations Area Files.
- BURTON, ROGER V.; and WHITING, JOHN W. M. 1961 The Absent Father and Cross-sex Identity. *Merrill-Palmer Quarterly of Behavior and Development* 7:85-95.
- CHILD, IRVIN L. 1954 Socialization. Volume 2, pages 655-692 in Gardner Lindzey (editor), *Handbook of Social Psychology*. Cambridge, Mass.: Addison-Wesley.
- CHILD, IRVIN L.; STORM, THOMAS; and VEROFF, JOSEPH 1958 Achievement Themes in Folk Tales Related to Socialization Practice. Pages 479-492 in John W. Atkinson (editor), *Motives in Fantasy, Action, and Society: A Method of Assessment and Study*. Princeton, N.J.: Van Nostrand.
- DOLLARD, JOHN (1935) 1949 *Criteria for the Life History: With Analyses of Six Notable Documents*. Gloucester, Mass.: Smith. → See especially Chapter 1.
- DUBOIS, CORA A. (1944) 1960 *The People of Alor: A Social-psychological Study of an East-Indian Island*. Cambridge, Mass.: Harvard Univ. Press. → A two-volume paperback edition was published in 1961 by Harper.
- DYK, WALTER 1938 *Son of Old Man Hat: A Navaho Autobiography*. New York: Harcourt.
- ERIKSON, ERIK H. (1950) 1964 *Childhood and Society*. 2d ed., rev. & enl. New York: Norton.
- FORD, CLELLAN 1941 *Smoke From Their Fires: The Life of a Kwakiutl Chief*. New Haven: Yale Univ. Press.
- FREUD, SIGMUND (1913) 1959 Totem and Taboo. Volume 13, pages ix-162 in Sigmund Freud, *The Standard Edition of the Complete Psychological Works of Sigmund Freud*. London: Hogarth; New York: Macmillan.
- GLADWIN, THOMAS 1961 Oceania. Pages 135-171 in Francis L. K. Hsu (editor), *Psychological Anthropology: Approaches to Culture and Personality*. Homewood, Ill.: Dorsey.
- GLADWIN, THOMAS; and SARASON, SEYMOUR B. 1953 *Truk: Man in Paradise*. Viking Fund Publications in Anthropology, No. 20. New York: Wenner-Gren Foundation for Anthropological Research.
- HERSKOVITS, MELVILLE J. 1948 *Man and His Works: The Science of Cultural Anthropology*. New York: Knopf.
- JOSEPH, ALICE; SPICER, ROSAMUND B.; and CHESKY, JANE 1949 *The Desert People: A Study of the Papago Indians*. Univ. of Chicago Press.
- KARDINER, ABRAM (1939) 1955 *The Individual and His Society: The Psychodynamics of Primitive Social Organization*. New York: Columbia Univ. Press.
- KARDINER, ABRAM 1945 *The Psychological Frontiers of Society*. New York: Columbia Univ. Press. → A paperback edition was published in 1963.
- KLUCKHOHN, CLYDE 1939 Theoretical Bases for an Empirical Method of Studying the Acquisition of Culture by Individuals. *Man* 39:98-105.
- KROEBER, ALFRED L. 1920 Totem and Taboo: An Ethnologic Psychoanalysis. *American Anthropologist* New Series 22:48-55.
- LAMBERT, WILLIAM W.; TRIANDIS, LEIGH M.; and WOLF, MARGERY 1959 Some Correlates of Beliefs in the Malevolence and Benevolence of Supernatural Beings: A Cross Societal Study. *Journal of Abnormal and Social Psychology* 58:162-169.
- LANDY, DAVID 1959 *Tropical Childhood: Cultural Transmission and Learning in a Rural Puerto Rican Village*. Chapel Hill: Univ. of North Carolina Press.
- LEIGHTON, DOROTHEA (CROSS); and KLUCKHOHN, CLYDE 1947 *Children of the People: The Navaho Individual and His Development*. Cambridge, Mass.: Harvard Univ. Press.
- MCCLELLAND, DAVID C.; and FRIEDMAN, G. A. 1952 A Cross-cultural Study of the Relationship Between Child-training Practices and Achievement Motivation Appearing in Folk Tales. Pages 243-249 in Society for the Psychological Study of Social Issues, *Readings in Social Psychology*. 2d ed. New York: Holt.
- MACGREGOR, GORDON 1946 *Warriors Without Weapons: Study of the Society and Personality Development of the Pine Ridge Sioux*. Univ. of Chicago Press.
- MALINOWSKI, BRONISLAW (1927) 1953 *Sex and Repression in Savage Society*. London: Routledge; New York: Harcourt. → A paperback edition was published in 1955 by Meridian.
- MEAD, MARGARET (1928) 1961 *Coming of Age in Samoa: A Psychological Study of Primitive Youth for Western Civilization*. New York: Morrow.
- MEAD, MARGARET (1930) 1953 *Growing Up in New Guinea*. New York: New American Library.
- MEAD, MARGARET (1935) 1950 *Sex and Temperament in Three Primitive Societies*. New York: Mentor.
- MEAD, MARGARET 1949 *Male and Female: A Study of the Sexes in a Changing World*. London: Gollancz; New York: Morrow. → A paperback edition was published in 1955 by New American Library.
- MINTURN, LEIGH; and LAMBERT, WILLIAM W. 1964 *Mothers of Six Cultures: Antecedents of Child Rearing*. New York: Wiley.
- PROTHRO, EDWIN T. 1961 *Child Rearing in the Lebanon*. Harvard University Middle Eastern Monograph No. 8. Cambridge, Mass.: Harvard Univ. Press. → Copyright 1961 by the President and Fellows of Harvard College. Extracts from this book are reprinted by the permission of the publishers.
- ROBERTS, J. M.; and SUTTON-SMITH, B. 1962 Child Training and Game Involvement. *Ethnology* 1:166-185.
- SAPIR, EDWARD (1917) 1949 Excerpt From a [Book] Review of Oskar Pfister, *The Psychoanalytic Method*. Pages 522-525 in Edward Sapir, *Selected Writings of Edward Sapir in Language, Culture and Personality*. Berkeley and Los Angeles: Univ. of California Press.
- SEARS, ROBERT R.; MACCOBY, E. E.; and LEVIN, H. 1957 *Patterns of Child Rearing*. Evanston, Ill.: Row, Peterson.
- SPIRO, MELFORD E.; and D'ANDRADE, ROY G. 1958 A Cross-cultural Study of Some Supernatural Beliefs. *American Anthropologist* New Series 60:456-466.
- TALAYESVA, DON C. 1942 *Sun Chief: The Autobiography of a Hopi Indian*. Edited by Leo Simmons. New Haven: Yale Univ. Press.
- THOMPSON, LAURA; and JOSEPH, ALICE (1944) 1947 *The Hopi Way*. Univ. of Chicago Press.
- TYLOR, EDWARD B. (1871) 1958 *Primitive Culture: Researches Into the Development of Mythology, Philosophy, Religion, Art and Custom*. Volume 1: Origins of Culture. Gloucester, Mass.: Smith.

- WHITING, BEATRICE B. (editor) 1963 *Six Cultures: Studies of Child Rearing*. New York: Wiley.
- WHITING, JOHN W. M. 1941 *Becoming a Kwoma: Teaching and Learning in a New Guinea Tribe*. New Haven: Yale Univ. Press.
- WHITING, JOHN W. M. 1954 The Cross-cultural Method. Volume 1, pages 523-531 in Gardner Lindzey (editor), *Handbook of Social Psychology*. Cambridge, Mass.: Addison-Wesley.
- WHITING, JOHN W. M. 1959a Sorcery, Sin, and the Superego: A Cross-cultural Study of Some Mechanisms of Social Control. Volume 7, pages 174-195 in Marshall R. Jones (editor), *Nebraska Symposium on Motivation*. Lincoln: Univ. of Nebraska Press.
- WHITING, JOHN W. M. 1959b Cultural and Sociological Influences on Development. Pages 5-9 in Maryland Child Growth and Development Institute, Baltimore, 1959, *Maryland Child Growth and Development Institute, June 1-5, 1959*. Baltimore: Maryland State Department of Health.
- WHITING, JOHN W. M. 1961 Socialization Process and Personality. Pages 355-380 in Francis L. K. Hau (editor), *Psychological Anthropology: Approaches to Culture and Personality*. Homewood, Ill.: Dorsey.
- WHITING, JOHN W. M.; and CHILD, IRVIN L. 1953 *Child Training and Personality: A Cross-cultural Study*. New Haven: Yale Univ. Press. → A paperback edition was published in 1962.
- WHITING, JOHN W. M.; KLUCKHOHN, RICHARD; and ANTHONY, ALBERT 1958 The Function of Male Initiation Ceremonies at Puberty. Pages 359-371 in Society for the Psychological Study of Social Issues, *Readings in Social Psychology*. 3d ed. New York: Holt.
- WHITING, JOHN W. M. et al. 1966a The Learning of Values. Pages 83-125 in Evon Z. Vogt and Ethel M. Albert (editors), *People of Rimrock: A Study of Values in Five Cultures*. Cambridge, Mass.: Harvard Univ. Press.
- WHITING, JOHN W. M. et al. 1966b *Field Guide for the Study of Socialization*. New York: Wiley.

### III

#### POLITICAL SOCIALIZATION

Narrowly conceived, political socialization is the deliberate inculcation of political information, values, and practices by instructional agents who have been formally charged with this responsibility. A broader conception would encompass all political learning, formal and informal, deliberate and unplanned, at every stage of the life cycle, including not only explicitly political learning but also nominally nonpolitical learning that affects political behavior, such as the learning of politically relevant social attitudes and the acquisition of politically relevant personality characteristics. For the present purposes we shall think of political socialization in the second, more comprehensive, sense.

#### The importance of political socialization

Political actions are determined both by the "objective" situations in which political actors find themselves (for example, the constraints and re-

wards imposed by the environment, including the structure of political institutions) and by the predispositions that citizens and their leaders have acquired through prior experience—*inter alia*, their political goals, expectations about the rules of the game of politics, conceptions of the legitimacy of men and institutions, group loyalties, assumptions about "human nature," and orientations toward authority. The nature of these predispositions and the processes through which they are acquired are matters of crucial political importance. For example, the political socialization process of a society may contribute to stability or instability, to continuity or change, to high or low levels of public political participation.

Socialization is an economical tool of government. To the degree that government relies upon the habituated responses of citizens, the necessity of environmental constraints is lessened. Authority is more stable when obeyed automatically than when sanctions must be threatened or employed. Yet it is not necessarily the stable, secure government that attends consciously and explicitly to political socialization. Rather, civic training becomes a deliberate policy when elements of potential instability are perceived in a political system—for instance, when an attempt is being made to weld together diverse and antagonistic populations into a single nation. Extensive formal political education also occurs where, as is the case under totalitarianism, the state carries on many of the functions ordinarily performed by other institutions, such as the family.

In the United States considerable explicit discussion of the importance of civic training can be found in the educational literature of the turn of the century, at which time sizable immigrant populations were being assimilated. Civic education is still considered a key function of the schools. Although detailed research has not been carried out, it seems likely that preoccupation with formal civic education and the inculcation of patriotism in the United States is greater than in Britain, but less than in an unstable democracy such as France, and certainly far less than in the Soviet Union, where "character training" and child-raising practices, as well as specifically political education, are officially discussed in terms of the requirements of the state (Bronfenbrenner 1962).

#### The study of political socialization

Scholarly interest in political socialization is coeval with the beginnings of political theory. It is difficult to think of a political theorist who exceeds Plato in his concern with the training of citizens.



[illegible]

These strategies of internationalisation starting after the war (1945) in the five states and their allies have been described traditionally as regional policy, not as global strategy. The formal documents of the planning were presented by a number of studies conducted by the American Historical Association (e.g. Nathan 1954, Flors 1961) and the meaning of 'regional' was the subject of a series edited by Philip H. Barbour (1962) in 1962. In 1962, the United States is said to pursue 'regional' development, 'consisting of efforts to promote self-sufficient and economic development. An important aspect consists of making significant advances in performance within each economic area, and to ensure that development efforts (Gardner & Lerner 1965: 101) [are] directed either towards either from the starting of their creation by the development of primary sources, especially studies of the political development of strategy and adjustment. During the period of the late post-war economic boom in the world in that region (Gardner 1966).

1. The present study was conducted in the following manner:

Although there is no primary interest in itself in the fairly important surrounding world, there is rather an inwardly self-regarding attitude, which is expressed in the following passage of *Lebensweisheit* (Knowledge of the Great Masters of Occultism), vol. III: "I regard what is mine more as mine than almost anything in which others are interested. I am selfish, I am egoistic, I am egomaniacal, and I am egomaniacal with a vengeance." (The translation will be somewhat more exact than the original, but the sense will be the same.)

For the purpose of this study, the following definitions were used. The term "social capital" was defined as the "resources of trust, reciprocity, and cooperation among individuals and organizations that facilitate collective action" (Putnam, 1993, p. 167). The term "social capital" was also defined as "the ability of a community to act together to solve problems" (Coleman, 1988, p. 30). The term "social capital" was also defined as "the ability of a community to act together to solve problems" (Coleman, 1988, p. 30).

1. The first step is to identify the problem or question that needs to be answered. This involves understanding the context and the specific requirements of the task.

upper social strata are more politically active than members of the lower strata. Research on children's political awareness, showing that these differences have already begun to be revealed by the age of age, indicates that there are no limits to social variation. Sex differences in political activity, too, seem to be encouraged by the entire psychological sex typing. Beginning with play activity, male children learn to be more sensitive to aspects of the environment beyond their personal needs and to be more active in their response to the environment. The political activity of young children, thus, reflects a socialization process from the young and male sex towards a greater awareness of the social and political environment. It is not clear, however, whether this socialization process is a function of the social and political environment or whether it is a function of the child's sex.

The volume is printed using University of Toronto's new "green" paper that is made from recycled material. The authors also express their appreciation for the support of the University of Toronto's Faculty of Education and the Ontario Ministry of Education. The authors also thank the many people who helped them with the book, including the many people who helped them with the book, including the many people who helped them with the book.

[illegible]

1973





agency for advancement of viewpoints related to party politics, but the schools are exposed to the general societal process of political and politically relevant communication and persuasion. Interest groups—industries, professions, liberal and conservative associations, patriotic societies, ethnic groups, agencies of government—vie for a hearing in the classroom. Groups provide teachers with instructional materials, send speakers to schools, and take an interest in the content of textbooks. (These groups also, of course, have direct programs designed to convey their views to children and adults.) From the earliest grades, schools purvey the symbols of nationalism through observance of holidays, teaching about national heroes, and patriotic ceremonies.

If the schools transmit communications from the entire range of societal institutions, so also do the mass media. The latter are extraordinarily important agents of political socialization. In the early 1960s the average American grade school child spent approximately twenty hours weekly viewing television and additional hours being exposed to other media (Witty et al. 1963). Many of the communications thus received are without specific political content, but some contact with political information is unavoidable. And children deliberately seek information about certain dramatic political events. The long-run effect of media attention is probably to build up, gradually and inadvertently, an awareness of basic elements in the political system.

There is much research suggesting that face-to-face communication in a primary group setting (family, peer group, neighborhood) is more persuasive than communication through the mass media. The study of political socialization supports this view strongly. In particular, the family is a vital source of socialization, especially in the acquisition of beliefs and loyalties. We have seen the apparent significance for French politics of the failure of families to serve as agencies of political communication. In the United States there is a high correlation between the political beliefs, especially the party preferences, of parents and those of their children: children acquire either partisan loyalties or the tendency not to have partisan loyalties from their parents (Hyman 1959; Greenstein 1965, chapter 4). Parents also have a powerful effect upon their children's motivation to participate in politics.

**Circumstances of political socialization.** All of the principles governing the effectiveness of educational practices and, more generally, of communication and persuasion, apply to assessing the effects of the various circumstances under which

political socialization takes place. A circumstance of political socialization alluded to above is the age at which learning takes place, including the sequence of learning.

That early learning is of decisive importance is suggested by many contemporary psychological theories and by traditional folklore. But the effects of the early formation of politically relevant personality structures may be attenuated or obliterated by later experience. Almond and Verba (1963) find a stronger correlation between disposition to participate in politics and feelings that one is free to participate at his work place than with feelings that earlier in life one was free to participate in the classroom or the home.

In the case of specifically political learning (in contrast to politically relevant personality development), there is evidence for the considerable importance of early experience. Political orientations learned during the initial school years or the late preschool years often have a greater impact on the individual's adult political behavior than do orientations that are learned later in life. For example, party identification is formed early in life; ideological orientations come much later. In the United States at least, the voter's party preference has a stronger effect on his electoral decision making than do his views about political issues. Children learn earlier about political executives than about legislative bodies. Among adult citizens there is more awareness of the former than of the latter and probably a tendency to attribute greater importance to executives such as the president, governor and mayor than to their legislative counterparts.

One aspect of this is the learning sequence. The concepts, information, and feelings that are first acquired serve as "filters" through which later perceptions must pass. Thus the child who "knows" that he is a Democrat, without attributing any substantive meaning to his party preference, will learn to orient himself toward the utterances of Democratic politicians and to acquire issue commitments consistent with his initial loyalty. The child who first becomes aware of a political executive and later learns of the existence of a coordinate legislative body may tacitly assume that the legislature is in some way ancillary or subordinate to the executive. Apart from its sequential effects, early learning may be important in and of itself because it takes place at a plastic, formative stage of development when critical capacities are limited. Adult orientations that have their roots in early learning and that have therefore been adopted without conscious consideration of alternatives, are likely to have an unquestioned character that

makes them both influential for behavior and resistant to change.

The effects of political socialization. Political learning has effects on the later political behavior of the individual exposed to socializing influences and, by extension, upon the political system. Most commonly, socialization seems to have conservative consequences for existing political arrangements. We have seen, for example, that class and sex differences in political participation tend to be transmitted from generation to generation. The conservative effects of socialization are not necessarily in the direction of encouraging political stability; political socialization in both stable and unstable societies is likely to maintain existing patterns.

Nevertheless, socialization is a potential source of change. It is always possible that the link will be broken in the attempt to transmit one generation's predispositions to the next. And since the training of the young is in part future-oriented, one generation may deliberately inculcate the next with values that differ from its own (Inkeles 1955). More generally, wherever different generations are exposed to different experiences, the seeds of change are present.

FRED I. GREENSTEIN

[Directly related are the entries ACCESS TO POLITICS, CONSENSUS, article on THE STUDY OF CONSENSUS, GENERATIONS, article on POLITICAL GENERATIONS, IDENTIFICATION, POLITICAL; LOYALTY; POLITICAL EFFICACY; POLITICAL RECRUITMENT AND CAREERS. Other relevant material may be found in PERSONALITY, POLITICAL, POLITICAL CULTURE, POLITICAL PARTICIPATION, POLITICAL SOCIOLOGY.]

#### BIBLIOGRAPHY

- ADORNO, T. W. et al. 1950 *The Authoritarian Personality*. American Jewish Committee, Social Studies Series, No. 3. New York: Harper.
- ALMOND, GABRIEL A.; and VERBA, SIDNEY 1963 *The Civic Culture: Political Attitudes and Democracy in Five Nations*. Princeton Univ. Press.
- AMERICAN ACADEMY OF POLITICAL AND SOCIAL SCIENCE 1965 *Political Socialization: Its Role in the Political Process*. Edited by Roberta Sigel. Annals, Vol. 361. Philadelphia: The Academy.
- BROOKS-GUNN, URS 1962 *Soviet Methods of Character Education: Some Implications for Research*. *American Psychologist* 17:559-564.
- COLEMAN, JAMES S., editor 1965 *Education and Political Development*. Princeton Univ. Press.
- CONVERSER, PHILIP E.; and DUFRUX, GEORGES 1962 *Politicalization of the Electorate in France and the United States*. *Public Opinion Quarterly* 26:1-23.
- EASTON, DAVID, and HESS, ROBERT D. 1962 *The Child's Political World*. *Midwest Journal of Political Science* 6:229-246.
- GREENSTEIN, FRED I. 1965 *Children and Politics*. Yale Studies in Political Science, Vol. 12. New Haven: Yale Univ. Press.
- HYMAN, HERBERT H. 1950 *Political Socialization: A Study in the Psychology of Political Behavior*. Glencoe, Ill.: Free Press.
- INKELES, ALEX 1955 *Social Change and Social Character: The Role of Parental Mediation*. *Journal of Social Issues* 11, no. 2:12-23.
- INKELES, ALEX; and LIVINGSON, DANIEL J. 1954 *National Character: The Study of Modal Personality and Sociocultural Systems*. Volume 2, pages 977-1020 in Gardner Lindzey (editor), *Handbook of Social Psychology*. Cambridge, Mass.: Addison-Wesley.
- LEVINE, ROBERT A. 1960 *The Internalization of Political Values in Stateless Societies*. *Human Organization* 19:51-58.
- MERRIAM, CHARLES E. 1931 *The Making of Citizens: A Comparative Study of Methods of Civic Training*. Univ. of Chicago Press.
- MERRIAM, CHARLES E. 1934 *Civic Education in the United States*. Report of the Commission on the Social Studies, American Historical Association, Part 6. New York: Scribner.
- PIERCE, BESSIE L. 1933 *Citizens' Organizations and the Civic Training of Youth*. Report of the Commission on the Social Studies, American Historical Association, Part 3. New York: Scribner.
- WITTY, PAUL A. et al. 1963 *A Summary of Yearly Studies of Televiewing: 1949-1963*. *Elementary English* 40:590-597.

#### IV

#### ADULT SOCIALIZATION

Every society is faced with the task of socializing its children into the basic culture and, to varying degrees, of providing further socialization of these persons as they move into different statuses at different stages in the life cycle. In the simplest terms, one can say that through socialization the individual acquires the culture of his group or groups. This includes two main divisions of culture: the traditional positions, or statuses, in the society and the role behaviors associated with them.

The socialization that an individual receives in childhood cannot be fully adequate preparation for the tasks demanded of him in later years. As an individual matures, he moves through a sequence of statuses corresponding to different stages in the life cycle. In addition, his interpersonal environment may change because of geographic or social mobility with consequent demands for new kinds of behavior. Even though some of the expectations of society are relatively stable through the life cycle, many others change from one position to the next.

It would seem desirable from society's point of view to be able to socialize an individual in childhood so that he could successfully handle all of the roles he would confront in the future. Perhaps this is possible in a relatively unchanging society with little mobility where one could have foresight about an individual's path through the whole life cycle. But this orderly state of affairs usually can-



not be achieved; it can only be approximated to varying degrees, from one society to the next. Society can do no more than lay the groundwork for the necessary learning in later life, when the child will be confronted with the as yet only dimly seen adult roles.

There are, to be sure, other reasons why childhood socialization may be ineffective in later years. One major cause is that the demands for behavior at different stages of the life cycle may conflict (Benedict 1938). There are many other causes. In any given case, the individual himself may be unable to learn the necessary skills. Again, there may be agents missing, as in the absence of a parent or key institutions or agencies. The process can also fail because subgroups with deviant values exist in every society, and they do not prepare the child for the performance of the roles expected of him by the larger society at a later date. Finally, the specific socializing agent, such as a parent, to whom socialization is entrusted by society, may be inadequate to carry out the task because he himself is not interested or because he is ignorant or emotionally disturbed.

**Adult socialization.** For analytic purposes, a typology of adult socialization can be constructed in such a way as to display the origins of the need for each type.

Consider first the individual who is confronted with a new role and knows virtually nothing about what he should do. In such a case society will require new socialization, two types of which can be distinguished. The first of these is socialization which is *legitimate*, in the sense that it is recognized as needed by society and the individual is not expected to have learned the role earlier. This is true of individuals moving through a sequence of student roles, and of job training of certain kinds. The second type can be referred to as *illegitimate*, in the sense that the individual should have learned the role earlier. Examples here come with greatest frequency from the marital and parental roles: deficiencies which may have been caused by inattention to the person's socialization for this role on the part of his parents, or by a variety of other influences on the individual's early development. Where the need for new socialization is recognized as legitimate, one usually finds formal institutional mechanisms in society to provide it. These include schools, vocational training programs, and similar devices. Where the need is not legitimate, retraining institutions do not usually exist, although the fact that they do exist in some areas, such as family counseling programs, reflects a growing recognition that the individual is not always to be blamed for ignorance of this kind of role.

The other major class of adult socialization is concerned with *resocialization*. Here the individual knows something about the role in question, but what he knows is wrong. For purposes of symmetry one can divide resocialization into two types along the dimension used above: the legitimacy or illegitimacy of the need for the resocialization. This time let us mention the illegitimate case first, since it is the more familiar one. Resocialization here includes rehabilitation of the criminal, handling the young delinquent through environmental manipulation, and therapeutic work with middle-class neurotics; in all three cases, resocialization is designed to make up for earlier socialization of an unsanctioned kind.

The legitimate type of resocialization—where the individual was socialized, but in the wrong way, so that the resocialization is viewed as legitimate and hence is sanctioned by society—is a less familiar instance. This is seen in the change from an ideal conception of a role to a more realistic conception, as illustrated in the relation of professional training to later professional practice (Becker et al. 1961) and, more generally, with respect to childhood myths about how adults behave, which clearly are functional in maintaining stability in the early years but need to be dispelled in later years, when the child actually takes on the role of adult.

### Studies of socialization

Most work in the study of personality development has told us little about how an individual develops his reciprocal, socially regulated interactions with other human beings or how he comes to understand role prescriptions and to distinguish between the important statuses in his society. There are historical reasons for this. Most work on socialization has come from the field of child development. But the emphasis from the beginning, in the child welfare stations established after 1920, has been more on maturation than socialization, more on development than learning; most of the output has consisted of studies of mental and physical development, and only to a much lesser extent of studies of the social and emotional aspects of development. The stimulus to much of the current work on socialization came from a different source: the work of Freud and related theories of personality. The effect of early life experiences on the development of personality traits that were believed to be fundamental and enduring characteristics of the individual was the initial focus of study. It was later, around 1930, that concepts emerging from cultural anthropology, especially cultural relativity and the plasticity of human nature, extended the

scope of studies of personality and led to a convergence of the interests mentioned above. The traditional work on child development, combined with concepts stemming from clinical theories of personality and enhanced by cross-cultural perspectives, resulted in some notable studies of socialization by Sears and his colleagues (1957), Whiting (1963), Levy (1943), and others. This work, for all of its contributions to our knowledge, still did not deal with role learning as the content of socialization. It is probably for this reason that it did not lead to a consideration of personality change in later life.

Analyses of the socialization of children in the family are numerous. Variations in child-care practices have been catalogued and analyzed by hundreds of investigators. Reviews in this area include those by Hoffman and Lippitt (1960), Clausen and Williams (1963), Bronfenbrenner (1958), Brim (1957), and several chapters in Hoffman and Hoffman (1964). The attempts to find important basic dimensions for the classification of such socialization practices have shown fairly consistently for more than a generation that practices can be efficiently described by two fundamental dimensions: positive affectivity as opposed to negative affectivity, and dominance as contrasted with permissiveness. One can note that analyses of socialization of children are less concerned now with actual training methods, such as education regarding cleanliness or home duties, than with the qualities of the interpersonal relationships. The supposition is that characteristics of relationships—for example, the two dimensions described above—are more important for socialization effectiveness than are the specific educational techniques.

Traditional work within school systems on the socialization process has involved studies of the teaching process. A recent comprehensive report is by Ryans (1960). The great number of such studies makes it impossible even to comment on them here. Over the past several years Lippitt and his colleagues (Lippitt & Van Egmond 1962) have reconceptualized the study of classroom socialization and have opened up new and promising areas of research that parallel areas of study in the family and other nonschool socialization settings.

Compared with the amount of information available about earlier age periods, we have little knowledge about the older age groups. There is a paucity of research on the techniques of socialization for the marital and parental roles. The few analyses of techniques used in occupations include the work of Howard S. Becker and his colleagues (Becker et al. 1961). The work on formal education of adults that has been most valuable has been in

family-life education. Most of the studies of family-life education are highly valuable analyses of how one deals with adult socialization, and the observations made in these studies regarding methods used and the causes of change in adult personality are applicable to other adult socialization processes. These studies show the need to use group interaction as the context for learning and group interchange as the most effective method. This follows from the facts that the adult is not a *tabula rasa* and that the educational problem is one of change, of erasing what exists and substituting something new. A detailed analysis of these adult education methods has been made specifically for parent education (Brim 1959).

One can say, in general, that the value of our knowledge of socialization techniques lies in what it shows to be undesirable—in the information it gives us about procedures that produce resistance, repression, resentment, escape, withdrawal, and habitual behavioral conformity. But we still are far from knowing which methods are more effective than others in achieving the goals of socialization. We may know more about techniques of instruction in the educational field than in other socialization contexts; but here, since the main requirement of the student is that he acquire information rather than that he learn attitudes and motives, the problem may be simpler. In regard to the family, we seem only now to be learning some of the ways in which socialization procedures bring about a greater internalization of parental values. At the adult level the number of good evaluation studies are few in number and come mostly from the field of family-life education and mental health counseling (see Brim 1959). Our understanding of the differential effectiveness of techniques is still cloudy; effectiveness seems to depend on the nature of the content being transmitted, the characteristics of the persons being socialized, the persons' relationship to the socializing agent, and similar factors.

Studies would make a valuable contribution if they related techniques and their consequences to more general theories of socialization and of personality, and thus led to a general understanding of socialization, rather than remaining just narrow studies of the "effectiveness" of educational techniques.

### Limits of later-life socialization

The limits of socialization in later life are set, on the one hand, by the biological capacities of an individual and, on the other hand, by the burden of earlier learning. The effectiveness of later-life socialization is a consequence of the interaction of



these two limits, with the added limit of the types of socialization methods available to society. These methods depend, in turn, primarily on the knowledge available about human behavior and, to a lesser extent, on technological developments.

**Biological limits.** By and large, a society's demands upon adults are tailored to the capacities of the average man; individuals who, for biological reasons, fall substantially below this average are usually debarred from natural progression through the life cycle and, hence, do not often become liable to later-life socialization. There are, however, two ways in which biological restrictions lead to disabilities and thus to limitations on later-life socialization. The first of these occurs primarily in an open-class society with a high level of achievement motivation. Upward mobility into ever more demanding roles may lead an individual to positions whose demands he is unable to meet because of limited intelligence or strength, or other biological attributes. The second occurs when war and other disasters destroy society's protection of the individual from the direct impact of nature, and the culture is no longer fitted to the average man. Thus individuals who are biologically adequate for the roles they will meet in the course of their normal life cycles may suddenly find themselves unable to live under the new and primitive conditions.

**Limits of earlier learning.** There are several reasons why the effects of early experience place important limits on later socialization. In the first place, attitudes learned in childhood are especially durable because they are continually being taught and just as continually reinforced (Brim 1960). Second, there is good reason to believe that during early socialization the bulk of the unconscious material of the personality is accumulated. The continuity of the individual personality (and probably its characteristic modes of defense as well) is therefore maintained by the inertia, so to speak, of unconscious forces relatively inaccessible to change by later socialization. Finally, it has been suggested (Caldwell 1962) that the human life cycle, like that of subhuman species, may contain critical periods at which human beings must learn certain things if they are to develop further. Failure to learn these things during the appropriate period may make it impossible for subsequent learning to take place.

Whatever the usefulness of these or other theories, both biological and learning limitations on later-life socialization certainly exist, although their exact nature is not yet understood. Nevertheless, the powerful arguments for the potent effects of early life experiences should not deter the study

of large and important changes that may take place in later life. The fact is that actual evidence on the effects of early life experience has, up to now, been mostly of a case history nature, coming from clinical practice. Too little attention has been given to the study of those important changes which may occur after childhood as a result of socialization experiences. A few important studies have begun to appear (Hathaway & Monachesi 1961; Kagan & Moss 1962; Reichard et al. 1962; Neugarten et al. 1964); it is to be hoped that many more will follow.

**Relationship to the socializing agent.** The fact that childhood socialization is usually so much more effective than adult socialization can be explained partly in terms of the different types of relationships that typically obtain between the individual and the socializing agent or agency at different stages in the life cycle. The relationship between child and parent is a highly affective one; by contrast, the adult socialization context is likely to be far less charged with emotion—in Parsons' phrase (1951, pp. 59–61), it is characterized by "affective neutrality." Moreover, the parent socializing the child is likely to make a far more open and continual use of power, so that the child can hardly avoid realizing that it is the weaker party in the situation. Agents of adult socialization, on the other hand, typically appeal more to the reason and self-interest of the person being socialized, and use power only as a last resort.

There is at least one major consequence of this difference for the results of socialization: adult socialization limits itself, on the whole, to a concern with behavior rather than with motivation and values. In fact, it is less able to teach basic values and probably requires a relationship paralleling that of childhood to bring about equivalent basic value changes. This may indeed occur—an example is adult religious conversion, in which the submissive relationship and high affective interchange with the religious figure underlie the radical shift in the adult's value system (Frank 1961). Another example is the extreme one of prisoner-of-war camps. Work on "brainwashing" and the breaking down of resistance to enemy values shows the context to be one in which the captors use their extreme power in a deliberate manipulation of the whole range of affect—from rejection and hate, on the one hand, to support and positive sympathy on the other—thus bringing the prisoner into a position similar to that of a child with his parent. [See BRAINWASHING; INTERNMENT AND CUSTODY.]

It follows that if society is to undertake basic re-socialization of adults in respect to motives and

values, it has to institutionalize in some form the high power and affectivity relationship characteristic of childhood learning.

### Changes in socialization content

The substantive content of socialization differs, of course, in important ways at different stages of the life cycle and in different major social institutions. Since both the needs for and the limits to socialization vary by life-cycle stages, it is probable that the types of content vary accordingly. I propose that we distinguish five major types of change in socialization content, as follows.

Perhaps the most important change is the shift in content from a concern with values and motives to a concern with overt behavior. Society assumes that the adult knows the values to be pursued in different roles, that he wants to pursue them with the socially appropriate means, and that all that might remain to be done is to teach him what to do. Thus society is willing to spend much less time in retraining adult motivation and values than in training children; it is understood that teaching basic values and motivations is a necessary task of the institutions serving children, especially the family, and they are organized to carry out this task.

Why should this difference exist? Probably it stems directly from the limitations on later-life learning that make it impractical to attempt a thorough resocialization. It may be that the costs are too high and that it simply is not efficient, from society's point of view, to spend too much time on teaching an old dog new tricks. Perhaps only in the case where the need for a certain kind of manpower is very great and the question of efficiency becomes secondary to the need for personnel can an intensive and costly resocialization effort be made for adults.

Society has at least two major solutions to this problem of resocialization. One is anticipatory, with attention being given to the selection of candidates for an adult organization. In order to screen out those who do not have appropriate motives and values for the anticipated roles. This helps to assure that those who enter the organization will not present difficult problems to the socialization program. In this way adults probably do get sorted out, more or less, and placed in social situations where they fit best in terms of the values and motives learned in their early life socialization. Hobbs 1963. A second solution is that society may accept conforming behavior alone as evidence of satisfactory socialization and may forgo any concern with value systems. This entails risk, for if the social

system undergoes stress, then the conformity, since it is superficial anyway, may break down rapidly.

The second of the changes in socialization content might be described as a change from the acquisition of new material to a synthesis of old material. As a person moves through the life cycle, he accumulates an extensive repertoire of responses, both affective and behavioral. These are organized in terms of roles and, at a more specific level, in terms of episodes within a role. These responses can be detached from the contexts in which they have been learned and used, and joined with others in new combinations suitable as social behavior responsive to the new demands of adulthood. One can say, therefore, that the content acquired in adult socialization is not so much new material as it is the aggregation and synthesis of elements from a storehouse of already learned responses, with perhaps the addition of several fragments that are newly learned when they are necessary to complete the complex social act demanded in a given situation. Socialization in the later-life stages seems to emphasize the practice of new combinations of skills already acquired, rather than teaching wholly new complexes of responses.

The third change in the content of socialization is the transformation of idealism into realism. As the individual matures, the society demands that he become more realistic and lay aside his childish idealism. The change in the content of expectations involves distinctions between statuses. Early learning encompasses the formal status structure; later learning takes into account the actual status structure, which may often be informal and unavowed. One designates as cynical a person who doubts that the actual and the formal are divergent; on the other hand, we think of a person as naive if he does not make this distinction. In socialization the young child is not taught much about the informal system, thus, in early years he may believe that the actual and the formal are nearly identical. This serves to maintain and legitimize the formal status differentiations and to protect them from change. However, as the child matures, the realistic aspects of status differentiation also must be taught if the system is to work effectively.

Closely related is learning to distinguish between ideal role prescriptions and that which is actually expected of one in a role. Here, as in the case of status differentiations, the inculcation of ideal role prescriptions results in a desirable idealism that strengthens and perpetuates the ideals of the society. As the child matures, he learns to take his part in society in terms of the realistic expecta-



tions of others rather than in terms of conformity to ideal norms.

The fourth type of change in socialization content is to a greater concern with teaching the individual to mediate conflicting demands. As one moves through the life cycle, he is forced to develop methods of choosing between conflicting role prescriptions. The possible conflicts between the expectations of reference set members can be classified into two main types. First, there is *intrarole* conflict, in which the expectations for performance of two or more individuals, or of one individual with respect to different aspects of the role, are in conflict. In the first instance, a man's wife and his employer may differ in their expectations of his performance on the job; in the second, the wife may expect her husband to be both companion and taskmaster to his son. Second, there is *interrole* conflict, which can also be classified into two subtypes: where the conflict is between two or more individuals with respect to two separate roles, as when the employer's demands for job performance conflict with the wife's demands for familial performance; and where the conflict is between the expectations of one individual for performance in two different roles, as when the wife has conflicting expectations for her husband's behavior at home and on the job.

The need to learn how to handle such conflicts occurs to a greater extent in later life for at least two reasons. First, if the cultural norm is that children should be protected from seeing life's conflicts, then it follows that nothing will be taught about ways of mediating them. Second, in later life there are more roles, as well as more complexity within roles, so that there is a much greater possibility of role conflict.

Thus, as one ages, he learns the ways of conflict resolution, which Ralph Linton (1945) once described so well: avoiding the situation, withdrawing acceptably from conflict, and scheduling conflicting demands in temporal sequence, so that the conflict disappears. In addition, one learns another very important method of conflict resolution, which I think has been overlooked. The fact is that in every society there are recognized prescriptions for solving certain kinds of conflicts that arise from the competing demands of reference set members. These prescriptions, which have been called "metaprescriptions" (Brim & Wheeler 1966), govern the resolution of conflict between demands on one's time and loyalties, and they usually, although not always, pertain to interrole rather than intrarole conflict. Examples of metaprescriptions are "Do what your employer asks of you, even if it means

that you have little time for your children," or "Side with your wife when she disciplines the children, even if you think she is wrong." It seems to me that a noticeable change in the content of socialization in later-life periods is the attention given to ways of resolving conflict through such metaprescriptions.

The fifth characteristic of change in socialization content is in the dimension of generality-specificity. In the context of the current discussion, this means that what is taught in socialization may apply either to many social situations or to just a few. The dimension of generality versus specificity can be applied to both components of role prescriptions, that is, to both values and means.

As a child, the individual is trained, both deliberately and unwittingly, by socializing agents in the goals and behavior appropriate for his sex. There are male and female styles of doing many different things, and these are learned early. Society tries to motivate the child to perform the behavior and to pursue the values expected of him, and it trains him in the necessary skills. These characteristics are general, in the sense that they are required in a variety of situations he will confront in society, either as major components or as necessary coloring of other aspects of his behavior.

The case is similar for cultural differences in basic values, such as those related to achievement, to other persons, to nature and the family, and indeed to all of those general value orientations, to use Florence Kluckhohn's phrase (1948), that help to distinguish between major cultural groups. They are acquired early (and, in contrast with sex roles, with perhaps less deliberate instruction), and they give shape and tone to the performance of many roles in society.

The individual is socialized for his socioeconomic position, or the style of life of a certain status level. In other words, he acquires general skills and values, appropriate to carrying out in a certain manner a number of specific role demands for behavior. The values and behavior characteristic of a subcultural group are usually acquired in childhood (C. Kluckhohn & Murray 1948; Miller & Swanson 1960; F. Kluckhohn & Strodtbeck 1961), and, as with sex roles and basic cultural values, some part of what is learned is gained outside of any deliberate formal training program. Socialization into a new social level or style of life occurs in later life also; for example, one function of certain college organizations, such as fraternities, is to carry on this kind of socialization into a social class level higher than that of the individual's family or origin. The existence of socializing agencies with

this function can be seen as a response to the legitimate need for resocialization that results from the upward social mobility in American life. [See REFERENCE GROUPS.]

### Deviance and social control

Deviance can be defined as failure to conform to the expectations of other persons. Since there are always groups with different viewpoints in society, what is regarded as deviant from one person's frame of reference might be considered conformity by another. Thus, whether an individual's behavior or values are deviant always must be determined by reference to the viewpoint of some particular person or group.

Deviance has its cause in an individual's ignorance, or his lack of ability, or his lack of motivation; and it may occur in behavior, or values, or both. We can recognize two major underlying causes of ignorance, inability, and lack of motivation (excluding biological limitations to effective socialization, such as low intelligence). One major cause is ineffective socialization of the individual for the performance of the roles expected of him, even though the social systems in which he lives have been relatively unchanging. The second major cause is a shift in what is expected of an individual resulting from social change and leaving the individual in a situation where his prior socialization, although quite adequate for performance with reference to the old role prescriptions, no longer serves him well.

The modes of attempted control over deviance reflect a society's predominant theories and assumptions about the causes of deviance and are rooted in basic ideas about human nature: for example, whether man is inherently a stupid animal, whether he is possessed by demons or controlled by other supernatural forces, whether he is innately depraved and burdened with original sin, and so on.

In our own society it is deviance in motivation and values that is viewed as most serious. The concept of motivation plays an important role in our theories of why human beings act as they do, and deviance in motivation is viewed as a serious threat to the social order. Thus there is a tendency to examine instances of deviance for possible motivational components in order to appraise how serious the deviance is. In return, the deviant person, challenged as he is to account for his behavior and faced with punishments for motivational deviance, which are customarily greater than for deviance of other types, will plead ignorance or lack of ability as cause of his actions. The result is that considerable time is spent, in both the courts of

law and the systems of informal social control, searching for possible motivational deviance behind the façade of ignorance or lack of ability. As an answer to the difficult problem of identifying the motivational component in deviance, we often seem to begin with the assumption that the cause is in fact motivational. Perhaps another reason for this a priori assumption is that it places the "blame" on the individual for his behavior, rather than upon society.

The burden of proof thus is placed upon the actor to show that his motives are pure. The demand that he do so is legitimate from society's viewpoint, since it is difficult to distinguish the actual existence of ignorance or lack of ability from hypocritical claims that one did not know what the rules were or that he was unable to live up to them. However, there is a price for using this approach to the resocialization of deviant persons. The treatment of deviance would be more effective if it made use of techniques that accord with the causes of the behavior: education, where ignorance is the cause; improved training, where lack of ability is the reason; and, where motivation is the problem, a planned and deliberately executed program of manipulation of rewards and punishments that would reorient the individual to the appropriate goals and behavior. If deviance comes from ignorance or lack of ability, and yet punishment is given in the mistaken idea that motivation is the cause, then a frequent result is the individual's rejection of the values of society that he formerly accepted.

The concern with motivation and the burden of proof for motivational purity is less in the early stages of the life span. Once again, this is true both of the informal family systems and, at least in American society, of the courts of law. Probably the reason is that time remains to train the child, and the socializing agencies remain in firm control of adequate rewards and punishments to influence the course of a child's interests. With each advancing year, however, instances of deviant motivation are viewed with more seriousness, and the child's responsibility for motivational conformity increases in accord with the age-graded developmental schemes that are accepted in his culture, until the full-scale responsibility of adulthood is demanded of him.

ORVILLE G. BRIM, JR.

[See also **DEViant BEHAVIOR**; **ROLE**; **SOCIAL CONTROL**; **STATUS, SOCIAL**.]

### BIBLIOGRAPHY

BECKER, HOWARD S. et al. 1961 *Boys in White: Student Culture in Medical School*. Univ. of Chicago Press.



- BECKER, WESLEY C. 1964 Consequences of Different Kinds of Parental Discipline. Pages 169-208 in Martin L. Hoffman and Lois W. Hoffman (editors), *Review of Child Development Research*. New York: Russell Sage Foundation.
- BENEDICT, RUTH (1938) 1949 Continuities and Discontinuities in Cultural Conditioning. Pages 297-308 in Patrick Mullahy (editor), *A Study of Interpersonal Relations*. New York: Heritage. → First published in Volume 1 of *Psychiatry*.
- BRIM, ORVILLE G. JR. 1957 The Parent-Child Relation as a Social System: I. Parent and Child Roles. *Child Development* 28:345-364.
- BRIM, ORVILLE G. JR. 1958 Family Structure and Sex Role Learning by Children: A Further Analysis of Helen Koch's Data. *Sociometry* 21:1-16.
- BRIM, ORVILLE G. JR. 1959 *Education for Child Rearing*. New York: Russell Sage Foundation.
- BRIM, ORVILLE G. JR. 1960 Personality Development as Role-learning. Pages 127-159 in Ira Iscoe and Harold Stevenson (editors), *Personality Development in Children*. Austin: Univ. of Texas Press.
- BRIM, ORVILLE G. JR.; and WHEELER, STANTON 1966 *Socialization After Childhood: Two Essays*. New York: Wiley.
- BRIM, ORVILLE G. JR. et al. 1962 *Personality and Decision Processes: Studies in the Social Psychology of Thinking*. Stanford Univ. Press.
- BRONFENBRENNER, URIZ 1958 Socialization and Social Class Through Time and Space. Pages 400-425 in Society for the Psychological Study of Social Issues, *Readings in Social Psychology*. 3d ed. New York: Holt.
- CALDWELL, BETTYE M. 1962 The Usefulness of the Critical Period Hypothesis in the Study of Filial Behavior. *Merrill-Palmer Quarterly of Behavior and Development* 8:229-242.
- CLAUSEN, JOHN A.; and WILLIAMS, JUDITH R. 1963 Sociological Correlates of Child Behavior. Part 1, pages 62-107 in National Society for the Study of Education, *Child Psychology*. Univ. of Chicago Press.
- DOLLARD, JOHN; and MILLER, NEAL E. 1950 *Personality and Psychotherapy: An Analysis in Terms of Learning, Thinking, and Culture*. New York: McGraw-Hill. → A paperback edition was published in 1965.
- FRANK, JEROME D. 1961 *Persuasion and Healing: A Comparative Study of Psychotherapy*. Baltimore: Johns Hopkins Press.
- GLICK, PAUL C. 1957 *American Families*. New York: Wiley.
- HATHAWAY, STARKE R.; and MONACHESE, ELIO D. 1961 *An Atlas of Juvenile MMPI Profiles*. Minneapolis: Univ. of Minnesota Press.
- HOBBS, NICHOLAS 1963 A Psychologist in the Peace Corps. *American Psychologist* 18:47-55.
- HOFFMAN, LOIS W.; and LIPPITT, RONALD 1960 The Measurement of Family Life Variables. Pages 945-1013 in Paul Mussen (editor), *Handbook of Research Methods in Child Development*. New York: Wiley.
- HOFFMAN, MARTIN L.; and HOFFMAN, LOIS W. (editors) 1964 *Review of Child Development Research*. Vol. 1. New York: Russell Sage Foundation.
- KAGAN, JEROME; and MOSS, HOWARD A. 1962 *Birth to Maturity: A Study in Psychological Development*. New York: Wiley.
- KLUCKHOHN, CLYDE; and MURRAY, HENRY A. (editors) (1948) 1953 *Personality in Nature, Society, and Culture*. 2d ed., rev. & enl. New York: Knopf.
- KLUCKHOHN, FLORENCE R. (1948) 1953 Dominant and Variant Value Orientation. Pages 342-357 in Clyde Kluckhohn and Henry A. Murray (editors), *Personality in Nature, Society, and Culture*. New York: Knopf.
- KLUCKHOHN, FLORENCE R.; and STRODTBECK, FRED L. 1961 *Variations in Value Orientations*. Evanston, Ill.: Row, Peterson.
- LEVY, DAVID M. 1943 *Maternal Overprotection*. New York: Columbia Univ. Press.
- LINTON, RALPH 1945 *The Cultural Background of Personality*. New York: Appleton.
- LIPPITT, RONALD; and VAN EGMOND, ELMER (editors) 1962 *Inventory of Classroom Study Tools for Understanding and Improving Classroom Learning Processes*. Ann Arbor: Univ. of Michigan, Institute for Social Research.
- MILLER, DANIEL R.; and SWANSON, GUY E. 1960 *Inner Conflict and Defense*. New York: Holt.
- NEUGARTEN, BERNICE L. et al. 1964 *Personality in Middle and Late Life*. New York: Atherton.
- PARSONS, TALCOTT 1951 *The Social System*. Glencoe, Ill.: Free Press.
- PARSONS, TALCOTT 1955 Family Structure and the Socialization of the Child. Pages 35-131 in Talcott Parsons and Robert F. Bales (editors), *Family, Socialization and Interaction Process*. Glencoe, Ill.: Free Press.
- REICHARD, SUZANNE K.; PETERSEN, PAUL G.; and LIVSON, FLORINE 1962 *Aging and Personality: A Study of Eighty-seven Older Men*. New York: Wiley.
- RYANS, DAVID G. 1960 *Characteristics of Teachers: Their Description, Comparison, and Appraisal*. Washington: American Council on Education.
- SEARS, ROBERT R.; MACCOBY, E. E.; and LEVIN, H. 1957 *Patterns of Child Rearing*. Evanston, Ill.: Row, Peterson.
- WHEELER, STANTON 1961 Socialization in Correctional Communities. *American Sociological Review* 26:697-712.
- WHITING, BEATRICE 1963 *Six Cultures: Studies of Child Rearing*. New York: Wiley.
- WRONG, DENNIS H. 1961 The Oversocialized Conception of Man in Modern Sociology. *American Sociological Review* 26:183-193.

## SOCIETAL ANALYSIS

The social sciences consist of methods, concepts, theories, and findings whose ultimate point of reference is human society in general. While most of their ideas and data have thus far been produced by students of Western societies, social scientists aspire to use these ideas and data as means of analyzing or illustrating more general phenomena. Since aspiration often outruns achievement, especially in man's attempts to understand himself, unwarranted generalization from Western experience is still common; but it is now generally recognized in principle—and both the increasing numbers of non-Western social scientists and the increasing concern with data drawn from non-Western socie-

ties illustrate this—that the enterprise is a common, panhuman one.

The inclusion in this encyclopedia of a series of articles on the societies of many (but not all) regions of the modern world represents a recognition of both the universality to which the social sciences aspire and the difficulty they experience in achieving it. If universality were easily achieved, such separate treatment would be unnecessary. The article on area studies recounts the recent efforts by universities, governments, and foundations to achieve a less parochial perspective [see AREA STUDIES]. The present article explores the nature of the difficulty to which these efforts are directed and describes the continuing role of the analysis of particular societies in social scientific inquiry.

### Universality and relativity

The universalistic, generalizing element in the social sciences is the legacy of the Enlightenment view of mankind as essentially unitary, a view which, significantly enough, had as one of its products an encyclopedia. From this perspective, what is characteristically human is what all men have in common: their capacity to shape their own futures through action guided by self-conscious thought, or, in the Enlightenment idiom, "reason." The social sciences themselves become the scientific disciplining of this self-direction in the pursuit of a converging future for all men. However, running alongside this universalistic, rationalistic view in postmedieval social thought, and often intertwined with it, is another vigorous and venerable standpoint which locates the distinctively human in the amazing variety of sociocultural worlds which human communities have created for themselves. If man is rational, he also lives within historically given sociocultural complexes which condition and direct his rationality. It thus becomes the task of social science to analyze the *diverse particular* societies and cultures within which human self-direction must operate.

These two views confronted each other most starkly in the great political debates which preceded and followed the French Revolution: for the Encyclopedists, particular societies and cultures were the creatures of oppressive governments and obscurantist churches which "natural" reason should seek to strip away in the interest of common "natural" rights; for Edmund Burke and other writers of the conservative reaction, rights and reason could find expression and meaning only within the "natural" framework of historically given religious and political institutions (Parkin 1956).

Political debates aside, the paradox is still with us in the form of the continuing tension in the social sciences between the notion of cultural relativity and that of the "psychic unity of mankind."

While the paradox has not been resolved—indeed, perhaps, cannot be resolved, since it is inherent in the subject matter of the social sciences—it is now somewhat better understood. Studies of man's biology have demonstrated a common human ancestry and a common capacity to create culture—or at any rate have found it impossible to demonstrate the reverse (Dobzhansky 1962; Washburn 1963). Studies of language have demonstrated the common capacity of all known linguistic systems to mediate culture and to intertranslate (Hymes 1961). At the same time, our contemporary understanding of man's emergence from his primate ancestry suggests that the very freedom, relatively speaking, from genetically encoded patterning of behavior which makes possible his self-conscious self-direction renders him particularly dependent for behavioral patterning upon systems of shared public symbols—"cultures"—which are highly variable and are characterized by complex internal organization (Geertz 1962; 1965). Furthermore, since the social sciences themselves are systems of shared public symbols produced by men in society, the social scientific enterprise cannot entirely escape the cultural-historical relativity which conditions human social life in general (Mannheim 1929-1931). Both human social life and our attempts to understand it are thus products of a continuous interaction between man's captivity within distinctive cultural worlds and his capacity to manipulate, and in part transcend, such worlds.

The aspiration and capacity to transcend, both in social life and in our analyses of it, are particularly prominent today. While the many distinctive cultural traditions were far from mutually isolated in the past (McNeill 1963), today, it is generally agreed, their interaction has reached a level unique in human history. The major "civilized" traditions, and even most of the nonliterate "tribal" ones, are now in intimate and continuous contact through movements of persons (including social scientists) and through new means of long-distance communication. First, in the context of Western imperial and quasi-imperial domination, and more recently in that of a world system of formally sovereign but mutually dependent nation-states, a potentially common pool of technology, ideas, and organizational forms has come into existence. Whatever dangers and opportunities this may present to mankind as a whole, it confronts social scientists with the paradox of relativity and



universality in new and more poignant forms: How can the social sciences, still deeply conditioned by a preoccupation with Western societies and their historical experience, understand both the continuing cultural diversity and the seemingly convergent processes of change exhibited by a world community of societies that are increasingly interconnected ecologically? In particular, how are we to view the contemporary development of those societies outside the north Atlantic homeland of the social sciences, whose behavior and fate at present so engage the attention of both social scientists and men of affairs? How far are concepts that have been lifted out of Western folk sociology and put into the technical vocabulary of the social sciences—concepts such as family, political party, feudalism, and law—adequate for a truly comparative social science? What is the most promising strategy for developing more adequate ones?

One example may serve to illustrate the difficulty. The studies of occupational prestige conducted by the National Opinion Research Center in the United States have been repeated in many other countries, both Western and non-Western, with results which appear to be strikingly similar in the sense that the same occupations appear to have very similar relative prestige in all industrial and industrializing societies (Inkeles & Rossi 1956). However, there is much debate concerning the meaning of these findings. Do they mean that there is a system of social stratification characteristic of all industrial societies and that this structure has an internal logic of its own which operates independently of particular cultural settings? Or does the selection of occupation as the focus of prestige build into these studies an essentially Western stratification configuration, thus masking the differences in meaning which different cultures may give to similar organizational forms (Lipset & Bendix 1959, chapter 10; Fallers 1964, pp. 117–121)? No doubt the answer lies somewhere in between, but this is hardly an adequate conclusion, for the territory "in between" is vast. As Western social scientific techniques and concepts are increasingly extended to other societies, such problems may be expected to proliferate.

#### Comparative macrosociological perspective

One strategy for meeting some of the difficulties just described is the comparative macrosociological one. This approach is "holistic" in the sense that it takes seriously the interrelatedness of elements within sociocultural wholes. It need not involve assuming the impossible task of dealing exhaustively with all aspects of a sociocultural whole, nor

need it imply the view that the human social world is divisible into functionally discrete, self-sufficient systems (see Levy 1952, chapter 3). It does, however, require an attention to the larger sociocultural context of the particular elements under study and an attempt to conceptualize such contexts. The logic of such an analysis is more clinical than generalizing; instead of using the data of diverse particular societies to test a broad generalization, it brings together diverse general ideas in the attempt to understand particular societies. In the context of the social sciences as a whole, of course, it contributes to generalization because the application of general ideas to particular societies is a means of testing their adequacy. Even though its central subject matter may be one society, an analysis in this genre is always either implicitly or explicitly comparative because the general ideas of which it makes use have their origins in earlier attempts to understand other societies, or even the same society at an earlier period.

On purely scientific grounds, such an approach is supported by our appreciation of the consequences for particular institutions and collectivities of involvement in larger-scale units (Steward 1950). Apart from this, however, the organization of the contemporary world into populist nation-states and clusters of states confronts men everywhere with the practical necessity to think analytically about such large-scale social units. Africans and Europeans, Russians and Americans are urged by their governments to support policies toward one another as units—policies informed by understandings which, however adequate or inadequate, are inevitably macrosociological.

Of course, different styles of comparative macrosociological analysis are possible. One—the bolder—is represented by the various recent attempts by social scientists to leap out of history and to construct, in a burst of creative imagination, a universally applicable conceptual scheme for the comparative analysis of all societies (Parsons 1951; Levy 1952) or particular aspects of all societies, such as the political (Easton 1953). Another, less ambitious, style involves the comparison of particular societies or aspects of societies, case by case, in the effort to test more limited propositions (e.g., Bendix 1956). [*These and other differences in comparative studies are discussed in ANTHROPOLOGY, article on THE COMPARATIVE METHOD in ANTHROPOLOGY; POLITICS, COMPARATIVE; SOCIAL INSTITUTIONS, article on COMPARATIVE STUDY.*]

Both of these styles have contributed substantially to the development of a more adequate comparative perspective for the social sciences; which

of them one pursues depends upon how seriously one takes the relativity side of our paradox—how impressed one is by the cultural-historical distinctness and “given-ness” of particular societies and, hence, by the difficulty of arriving at universal categories. The regional articles in this encyclopedia, by their very nature, contribute to the less ambitious style in comparative macrosociology. These articles differ rather widely in focus and in analytical explicitness; but they all represent attempts to deal in a social scientific way with large-scale units, and they all employ intersocietal comparisons in an effort to increase our understanding of the societies under study. Often, either implicitly or explicitly, they employ concepts taken from Western thought about Western societies, and they thus raise the question of the explanatory power of such concepts in different cultural-historical settings. It may be suggested that explicitness in these matters is a virtue and that a self-conscious reference to the Western roots of our ideas and methods can contribute to the success of the comparative enterprise. Before discussing some of the propositions which emerge from the regional articles, therefore, it is appropriate to recall the historical context in which the macrosociological perspective emerged in Western thought.

**The emergence of the perspective.** The preoccupation with sociocultural differences is as old as the social sciences themselves, as Aristotle's comparative studies of constitutions and Herodotus' accounts of the institutions of non-Greek peoples show. One can, however, trace the immediate intellectual ancestry of contemporary comparative macrosociological concerns to more recent origins in the reflections of Western intellectuals upon their own past as they entered the postmedieval world of accelerating social change. There are, to be sure, gems of Western-non-Western comparison to be found in the literature of the Renaissance and the Enlightenment, such as Machiavelli's perceptive analysis of the French and Ottoman politics ([1532] 1941, pp. 104-105). For the most part, however, early modern intellectuals' interests were so seized by the revolutionary transformations occurring in their own societies that they used the data of non-Western societies (especially the non-literate societies of which they were being made aware by the voyages of discovery) more for the purpose of allegorical speculation upon the roots of their own institutions than for serious comparative analysis. It was the increasing awareness of change within Western societies and the growing knowledge of their ancient and medieval pasts produced by new historical scholarship that proved

most fruitful for comparative macrosociology until almost the end of the nineteenth century. The comparative perspective which emerged was one concerned with developmental stages in time rather than with differences and similarities among contemporary societies. However, as the Western societies achieved world dominance, especially in the nineteenth century, and as Western intellectuals became more conscious of their societies' preeminence, the comparative mode of thought was extended to the non-Western world, which was often seen as representing stages through which the West had passed. The Western experience came to be viewed as the end of history—a paradigm for understanding the experience of non-Western societies.

The classics of early comparative macrosociological thought were centrally concerned with characterizing and analyzing the new institutional complexes which set the societies of their day apart from those from which they had emerged. To mention only a few of the landmarks: Hobbes and Bodin sought to understand the nature of the new sovereign, secular state; Adam Smith analyzed the conditions for the politically autonomous market economy; Tocqueville considered the significance of increasing popular participation in politics; Marx traced the consequences of the separation of labor, under the system of industrial production, from the personalistic ties of kinship and feudal patronage; Weber considered the causes and consequences of the pervasive development of bureaucratic organization. From these studies emerged a set of conceptual models of the characteristic institutions of modern societies.

Other writers in this vein sought, through broad developmental typologies (“custom” and “reason,” “status” and “contract,” *Gemeinschaft* and *Gesellschaft*, “mechanical” and “organic”), to grasp intellectually the changing nature of whole societies. While there were occasional doubts among writers of this genre about the universal applicability of their developmental schemes—Marx, for example, considered the possibility of a distinctive “Asiatic mode of production” as an alternative to feudalism in precapitalist societies—on the whole, the question was either ignored or handled by assimilating non-Western forms as earlier stages. Nineteenth-century anthropology, while it contributed some classics of ethnographic description, was generally cast in a similar intellectual mold.

For some other Western intellectuals, the experience of radical social transformation stimulated thoughts of quite a different kind, both in relation to their own societies and, later, in relation to the









prominent role in both political and economic life. The author of the article on Japanese society concludes, *contra* Veblen (who in this instance represents the earlier view that the modernization of the West is paradigmatic), that this combination may be a relatively stable one and that it may be necessary to recognize a distinctively Japanese "modernism."

China is less fully modern in the economic sense, but the recent revolutionary transformation of Chinese society under an ideology drawn from one stream of Western experience poses an equal challenge to comparative understanding. As the article on Chinese society shows, a persistent problem for students of the Chinese revolution has been the question of whether the revolution is "essentially Marxist" or "essentially Chinese," i.e., whether it represents a basic change in society and culture along Marxist lines or a nationalist reassertion of traditional Chinese identity [see CHINESE SOCIETY]. The author acknowledges important elements of continuity but argues, in Weberian terms, that the emergence of a new dominant culture-bearing collectivity (the party) suggests a fundamental transformation. Incidentally, his discussion of the older dominant group, the "gentry," raises the question of the adequacy of this term, taken from English history, for understanding the premodern Chinese situation. The article on southeast Asia raises similar questions concerning the appropriate models for conceptualizing the structure of premodern societies in that area [see ASIAN SOCIETY, *article on SOUTHEAST ASIA*].

Several of the regional articles raise the question of the nature of the state in areas in which state and nation are in radical discontinuity. The model of the state assumed by the modern Western institutional paradigm is one in which a culturally and linguistically united people—a nation—is governed by a political institution possessing a monopoly of ultimate power within its boundaries and autonomy vis-à-vis other states (Kohn 1944; Jouvanel 1955). Although modern Western experience contains numerous exceptions to this model (Belgium and Switzerland have contrived to maintain states comprising culturally diverse groups, while as imperial powers several European states have exercised varying degrees of sovereignty over dependent polities, both within and without Europe), still the model has been felt to be paradigmatic in the sense that there is a strain toward sovereignty (decolonization) and, especially as populist politics develop, toward a coincidence between political and cultural boundaries (nationalism).

Today this conception is being tested in the ex-

perience of the new states and in social scientists' analyses of this experience. On the one hand, as the article on sub-Saharan Africa shows, the new states of that region—the new "global societies"—typically contain many formerly autonomous societies whose present political and cultural unity is mostly the result of European colonial rule [see AFRICAN SOCIETY, *article on SUB-SAHARAN AFRICA*]. On the other hand, the Arabs, whose present-day societies are analyzed in the article on north Africa and in one on the Near East, find their sense of linguistic, cultural, and religious unity violated by political fragmentation [see AFRICAN SOCIETY, *article on NORTH AFRICA*; NEAR EASTERN SOCIETY, *article on THE ISLAMIC COUNTRIES*]. Once united with non-Arab coreligionists within the mosaic empire of the Ottomans, they have achieved political independence but not political community. However "artificial" and "unstable" they may be, modern state boundaries in Africa and the Near East have proved highly resistant to change in response to ideological stirrings, suggesting that organizational forms may take on a life of their own that is independent, to some degree, of cultural forces.

Yet another pattern of state formation is illustrated (perhaps "constituted" would be the better term, in view of the uniqueness of the case) by Israel, a society formed through the recolonization of their homeland by a religious community dispersed for two thousand years throughout Europe and the Islamic world. Culturally and linguistically differentiated by the Diaspora, the Jews nevertheless retained sufficient sense of community, first to form a cohesive organizational structure across existing political boundaries and then, around this structure, to establish and defend a territorial base. [See NEAR EASTERN SOCIETY, *article on ISRAEL*.]

While the regional articles are far from supporting the notion that the present-day development of the non-Western world may be understood as simply a process of "Westernization," Western imperial influence has nevertheless been an important factor in the history of most societies outside Europe, and these articles therefore provide much material for the comparative study of imperialism. For the Old World non-European societies, imperialism represents an interlude—often far-reaching in its consequences but an interlude nevertheless—in histories in which the elements of continuity are indigenous. For the societies of the New World, however, the imperial imprint has been much more decisive. Even in Middle America and South America, where there were some large-scale indigenous societies with highly complex cultures, European

immigration was so massive and its cultural impact so profound that the resulting contemporary societies are more accurately viewed as American variants of Iberian society than as modernized or Westernized native American ones [see MIDDLE AMERICAN SOCIETY; SOUTH AMERICAN SOCIETY]. In the societies of the Caribbean, continuity from the indigenous past is still more tenuous. Populated largely by persons of African slave descent, these societies remain variously Spanish, English, and French in dominant language and culture and "colonial" in socioeconomic structure [see CARIBBEAN SOCIETY].

In the article on Anglo-American society we have a thoroughgoing comparative analysis of a European imperial society and its daughter societies overseas. Utilizing Parsons' scheme of pattern variables, the author analyzes the variations upon common sociocultural themes which arose in Great Britain, Canada, Australia, and the United States and concludes that today these differences are declining in response to common developmental trends. [See ANGLO-AMERICAN SOCIETY.]

Continental Europe, whose intellectuals have played such a central role in the development of the social sciences and whose societies have provided so much of their data, occupies a special place in comparative macrosociological studies. From one perspective—the one stressed here—its common sociocultural features have provided a paradigm for the comparative study of other societies. But while these common features are real enough, so that one might speak of the region as a "culture area" in the anthropological sense, it is (like other regions) internally differentiated in ways that invite comparative study on a more detailed level. This perspective is represented in this encyclopedia by the articles on various sociocultural complexes that are historically characteristic of Europe, such as democracy, socialism, communism, capitalism, Christianity, bureaucracy, political parties, and labor unions, all of which are analyzed comparatively. Indeed, much of the content of this encyclopedia, both substantive and conceptual-theoretical, is at least implicitly concerned with intra-Western comparative analysis. [See BUREAUCRACY; CAPITALISM; CHRISTIANITY; COMMUNISM; DEMOCRACY; LABOR UNIONS; PARTIES, POLITICAL; SOCIALISM.]

Of all the regional articles, the one on south Asian society gives least attention to the problem of cultural continuity and discontinuity [see ASIAN SOCIETY, *article on SOUTH ASIA*]. Acknowledging that Indic and Islamic cultures continue to inform the societies of the area, the author argues pro-

vocatively for an austere social structural analysis in terms of levels of complexity. This analysis of south Asian society also reminds us that models for comparative investigation are today drawn from diverse areas. While the author implicitly draws upon Western models for the analysis of the more complex levels of south Asian society, his frame of reference for analyzing village and tribal communities, he says, is derived from social anthropological studies of Africa. The study of south Asian society, in turn, has provided at least one model whose applicability to other societies social scientists have found it enlightening to explore. The article on the concept of caste, the area's most distinctive sociocultural complex, examines the south Asian phenomenon, compares it with others and, in a manner which well exemplifies the logic of comparative macrosociological analysis, reformulates it so as to elucidate intersocietal similarities and differences [see CASTE, *article on THE CONCEPT OF CASTE*].

The native societies of North America [see INDIANS, NORTH AMERICAN] and of Australia [see OCEANIAN SOCIETY] have been so displaced by European immigration that the article on Anglo-American society, which gives an account of the modern societies of these areas, can completely ignore them. The island societies, still under colonial rule, are perhaps destined for some sort of independence, a situation which, if it develops, will test some of the hypotheses suggested in the article on "small societies" [see SOCIETIES, SMALL]. While of little immediate relevance to the comparative study of contemporary national societies, research on these societies has been extremely productive of ideas which have influenced comparative macrosociological thought, for it was in relation to them that Malinowski, Boas, and Radcliffe-Brown developed the new anthropology, which emphasized holism. As the articles mentioned above show, continued anthropological research on these societies has produced a rich literature of comparative analysis, both synchronic and diachronic.

### The future

The scholarship which underlies the regional articles is the fruit of the great post-World War II expansion of the social sciences' concern with the non-Western, or "developing," world. Much of this activity has been carried out in the context of "area studies" programs, interdisciplinary programs of research and training organized around regions of the world characterized by common cultural traditions. In terms of the dialectical pattern of development in comparative macrosociological studies outlined above, the emphasis during this period has



been on the relativizing side: in an effort to make their social scientific concepts and theories more relevant to the task of understanding non-Western societies, many social scientists have become Indologists, Africanists, or China specialists as well as, or even in some cases instead of, social scientists. Recent developments suggest that today at least a partial reversal in emphasis may be in progress. As the results of specialized regional studies become more readily available to the social scientific community at large through secondary works of synthesis, somewhat more ambitious styles of comparative analysis again become profitable.

In the vanguard of this movement have been the students of economic development—the economists and economic historians, who have all along reached for generalization more persistently than other social scientists have. Few economists have become "area specialists." The reason for this is, of course, quite simple. From the time of Adam Smith to the present, the criterion of economic growth in monetary terms has provided economists with a relatively unambiguous standard about which to organize data for intersocietal comparison. Thus, while some, like Gerschenkron (1962), Hirschman (1963), and Rosovsky (1961), have become deeply engaged with the data of particular societies, their work has articulated directly with that of such generalizers as Bauer and Yamey (1957), Lewis (1955), and Schultz (1964).

For other social scientists, at any rate insofar as they have not been centrally concerned with the sociocultural prerequisites or concomitants of economic growth, there exists no such universal standard. Although a number of writers have attempted to conceptualize a "modernity" toward which all contemporary societies aspire or are tending (Shils 1959-1960; Kerr 1960). It is widely recognized that tradition remains a crucial element in all societies and that consequently there are likely to be many paths toward the future. Thus, for example, the analytic scheme proposed by Levy (1966), perhaps the most elaborate and rigorous in the recent literature of comparative macrosociology, is much more differentiated and open-ended than those of most nineteenth-century writers. The Western paradigm has been thoroughly relativized.

Levy's work is directed toward the comparative analysis of whole societies; much more common are comparative studies of particular aspects or institutions such as Janowitz's study of the military (1964), Apter's volume on politics (1965), and Lenski's work on social stratification (1966). Given the rising standards of regional scholarship, much

comparative work is, of necessity, collaborative. The volumes produced by the Committee on Comparative Politics of the Social Science Research Council (Almond & Coleman 1960; Conference on Communication . . . 1963; LaPalombara 1963; Conference on Political Modernization . . . 1964), the Committee for the Comparative Study of New Nations of the University of Chicago (Chicago . . . 1963; Johnson 1967), and the Congress for Cultural Freedom (*Democracy* . . . 1959; Bellah 1965) exemplify recent efforts to achieve comparative perspective through seminars and workshops of regional specialists.

Finally, as non-Western studies have matured, there has developed a discernible tendency toward completing the circle of comparative analysis by bringing their results to bear upon the history of the West. Lipset (1963) has re-examined the United States as "the first new nation," Moore (1966) has compared the historical role of agrarian organization in Europe and Asia, and Palmer (1959-1964), in his monumental study of the Western revolutions of the late eighteenth century, has felt it appropriate to comment upon the relationship between the events with which he is concerned and those of the twentieth century in the non-Western world. Bendix (1964) has compared the development of authority and citizenship in western Europe, Russia, India, and Japan, Eisenstadt (1966) has analyzed comparatively movements of protest and processes of change in the West and in the "third world." This article may therefore end with the observation with which it began: the road toward a more universal perspective for the social sciences leads through the study of diverse particular societies, considered comparatively.

LLOYD A. FALLERS

#### BIBLIOGRAPHY

- ALMOND, GABRIEL A.; and COLEMAN, JAMES S. (editors) 1960 *The Politics of the Developing Areas*. Princeton Univ. Press.
- APTER, DAVID E. 1965 *The Politics of Modernization*. Univ. of Chicago Press.
- BAUER, PETER T.; and YAMEY, BASIL S. 1957 *The Economics of Underdeveloped Countries*. Univ. of Chicago Press.
- BELLAH, ROBERT (editor) 1965 *Religion and Progress in Modern Asia*. New York: Free Press.
- BENDIX, REINHARD 1956 *Work and Authority in Industry: Ideologies of Management in the Course of Industrialization*. New York: Wiley. → A paperback edition was published in 1963 by Harper.
- BENDIX, REINHARD (1960) 1962 *Max Weber: An Intellectual Portrait*. Garden City, N.Y.: Doubleday.
- BENDIX, REINHARD 1964 *Nation-building and Citizenship. Studies of Our Changing Social Order*. New York: Wiley.

- CHICAGO, UNIVERSITY OF, COMMITTEE FOR THE COMPARATIVE STUDY OF THE NEW NATIONS 1963 *Old Societies and New States: The Quest for Modernity in Asia and Africa*. Edited by Clifford Geertz. New York: Free Press
- CONFERENCE ON COMMUNICATION AND POLITICAL DEVELOPMENT, DOBBS FERRY, N.Y., 1961 1963 *Communications and Political Development*. Edited by Lucian W. Pye. Princeton Univ. Press.
- CONFERENCE ON POLITICAL MODERNIZATION IN JAPAN AND TURKEY, GOULD HOUSE, 1962 1964 *Political Modernization in Japan and Turkey*. Edited by Robert E. Ward and Dankwart A. Rustow. Princeton Univ. Press
- DAHRENDORF, RALF (1957) 1959 *Class and Class Conflict in Industrial Society*. Rev. & enl. ed. Stanford Univ. Press. → First published as *Soziale Klassen und Klassenkonflikt in der industriellen Gesellschaft*
- Democracy in the New States: Rhodes Seminar Papers*. 1959 New Delhi: Congress for Cultural Freedom, Office for Asian Affairs.
- DOBZHANSKY, THEODOSIUS 1962 *Mankind Evolving: The Evolution of the Human Species*. New Haven: Yale Univ. Press.
- EASTON, DAVID 1953 *The Political System: An Inquiry Into the State of Political Science*. New York: Knopf.
- EISENSTADT, SHMUEL N. 1966 *Modernization: Protest and Change*. Englewood Cliffs, N.J.: Prentice-Hall
- FALLERS, LLOYD A. (editor) 1964 *The King's Men: Leadership and Status in Buganda on the Eve of Independence*. Oxford Univ. Press.
- GEERTZ, CLIFFORD 1962 *The Growth of Culture and the Evolution of Mind*. Pages 713-740 in Jordan M. Scher (editor), *Theories of the Mind*. New York: Free Press
- GEERTZ, CLIFFORD 1965 *The Impact of the Concept of Culture on the Concept of Man*. Pages 93-118 in John R. Platt (editor), *New Views of the Nature of Man*. Univ. of Chicago Press.
- GERSCHENKRON, ALEXANDER A. 1962 *Economic Backwardness in Historical Perspective: A Book of Essays*. Cambridge, Mass.: Harvard Univ. Press
- HIRSCHMAN, ALBERT O. 1963 *Journeys Toward Progress: Studies of Economic Policy-making in Latin America*. New York: Twentieth Century Fund
- HUBERT, RENÉ 1931 Encyclopédies Volume 5, pages 527-531 in *Encyclopaedia of the Social Sciences*. New York: Macmillan.
- HYMES, DELL 1961 *Functions of Speech: An Evolutionary Approach*. Pages 55-83 in Frederick C. Gruber (editor), *Anthropology and Education*. Philadelphia Univ. of Pennsylvania Press.
- INKELES, ALEX; and ROSSI, PETER H. 1956 *National Comparisons of Occupational Prestige*. *American Journal of Sociology* 61 329-339.
- JANOWITZ, MORRIS 1964 *The Military in the Political Development of New Nations: An Essay in Comparative Analysis*. Univ. of Chicago Press.
- JOHNSON, HARRY G. (editor) 1967 *Economic Nationalism in Old and New States*. Univ. of Chicago Press
- JOUVENEL, BERTRAND DE (1955) 1957 *Sovereignty. An Inquiry Into the Political Good*. Univ. of Chicago Press. → First published as *De la souveraineté: A la recherche du bien politique*.
- KERR, CLARK 1960 *Changing Social Structures*. Pages 348-359 in Wilbert E. Moore and Arnold S. Feldman (editors), *Labor Commitment and Social Change in Developing Areas*. New York: Social Science Research Council.
- KOHN, HANS 1944 *The Idea of Nationalism: A Study of Its Origins and Background*. New York: Macmillan. → A paperback edition was published in 1961.
- LAPALOMBARA, JOSEPH G. (editor) 1963 *Bureaucracy and Political Development*. Studies in Political Development, No. 2. Princeton Univ. Press.
- LENSKI, GERHARD 1966 *Power and Privilege: A Theory of Social Stratification*. New York: McGraw-Hill.
- LEVY, MARION J. JR. 1952 *The Structure of Society*. Princeton Univ. Press
- LEVY, MARION J. JR. 1966 *Modernization and the Structure of Society. A Setting for International Affairs*. 2 vols. Princeton Univ. Press
- LEWIS, W. ARTHUR 1955 *The Theory of Economic Growth*. Homewood, Ill.: Irwin; London: Allen & Unwin
- LIPSET, SEYMOUR M. 1963 *The First New Nation: The United States in Historical and Comparative Perspective*. New York: Basic Books.
- LIPSET, SEYMOUR M., and REYNOLDS, RICHARD 1959 *Social Mobility in Industrial Society*. Berkeley: Univ. of California Press
- MACHIAVELLI, NICCOLÒ (1532) 1941 *The Prince, and Other Works*. With an introduction and notes by Allen H. Gilbert. Chicago: Packard
- MCNEILL, WILLIAM H. 1963 *The Rise of the West: A History of the Human Community*. Univ. of Chicago Press
- MANNHEIM, KARL (1929-1931) 1954 *Ideology and Utopia: An Introduction to the Sociology of Knowledge*. New York: Harcourt, London: Routledge → A paperback edition was published in 1955 by Harcourt Part 1 is an introductory essay, Parts 2-4 are a translation of *Ideology and Utopia* (1929); Part 5 is a translation of the article "Wissenschaftslogie" (1931).
- MOORE, BARRINGTON JR. 1966 *Social Origins of Dictatorship and Democracy: Lord and Peasant in the Making of the Modern World*. Boston: Beacon.
- PALMER, ROBERT R. 1950-1964 *The Age of the Democratic Revolution: A Political History of Europe and America, 1760-1800* 2 vols. Princeton Univ. Press
- PARKIN, CHARLES 1956 *The Moral Basis of Burke's Political Thought*. Cambridge Univ. Press
- PARSONS, TALCOTT (1937) 1949 *The Structure of Social Action: A Study in Social Theory With Special Reference to a Group of Recent European Writers*. Glencoe, Ill.: Free Press
- PARSONS, TALCOTT 1951 *The Social System*. Glencoe, Ill.: Free Press
- ROBOVSKY, HENRY 1961 *Capital Formation in Japan, 1868-1940*. New York: Free Press
- SCHULTZ, THEODORE W. 1964 *Transforming Traditional Agriculture*. New Haven: Yale Univ. Press
- SKIDL, EDWARD (1950-1960) 1962 *Political Development in the New States. The Hague: Mouton*. → First published in Volume 2 of *Comparative Studies in Society and History*
- STEWART, JULIAN H. 1950 *Area Research. Theory and Practice*. Bulletin No. 63. New York: Social Science Research Council
- WASHBURN, S. L. 1963 *The Study of Race*. *American Anthropologist* New Series 65 531-531.
- WEBER, MAX (1914-1915, 1924) *The Protestant Ethic and the Spirit of Capitalism*. Translated by Talcott Parsons, with a foreword by R. H. Tawney. London: Allen & Unwin; New York: Scribner. → First published in German. The 1930 edition has been reprinted



frequently. A paperback edition was published in 1958 by Scribner.

WEBER, MAX (1906-1924) 1946 *From Max Weber: Essays in Sociology*. Translated and edited by Hans H. Gerth and C. Wright Mills. New York: Oxford Univ. Press. → First published in German.

WEBER, MAX (1922) 1957 *The Theory of Social and Economic Organization*. Edited by Talcott Parsons. Glencoe, Ill.: Free Press. → First published as Part 1 of *Wirtschaft und Gesellschaft*.

## SOCIETIES, SMALL

"Small" is obviously a relative term. When applied to societies it usually refers to area, population, or both; for purposes of sociological analysis a distinction must be drawn between a small-scale society and a small territory. It is possible to have a small-scale society in a very large territory (e.g., the Eskimo or the Bedouin). It is also possible to have part of a large-scale society in a very small territory (e.g., Luxembourg or Monaco). The criteria of size for territories are area and population; the criteria of scale for a society are the number and quality of role relationships. In a small-scale society the individual interacts over and over with the same individuals in virtually all social situations. In a large-scale society the individual has many impersonal or part-relationships. As Mair puts it, "Every member of the large-scale society is party to a great number of relationships, some ephemeral, some lasting, which do not overlap" (1963, p. 13).

Sociologists and social anthropologists have treated "smallness" in terms of the small groups found within a society. Usually this has been termed the "primary group" or the "face-to-face" group. Sometimes the general characteristics of such groups have been the focus of interest (e.g., Homans 1950); sometimes it has been the study of particular types of groups, such as the family (e.g., Bell & Vogel 1960), the gang (e.g., Whyte 1943), or some other form of association. Such face-to-face groups exist in all societies and are not a particular characteristic of small territories.

The study of "small-scale" societies has been the particular concern of social anthropologists. Indeed, smallness in scale has been cited as a distinguishing feature of the subject matter of social anthropology (Firth 1951, p. 17) and as a defining characteristic of "primitive" (Evans-Pritchard 1951a, p. 8). One should distinguish between two major types of small-scale societies. Both are composed chiefly of primary groups, but in one the total social field is small and in the other it is composed of a series of interlocking similar small groups which

extend through a considerable population. Island societies such as those on Tikopia (Firth 1936) or Dobu (Fortune 1932) are examples of the former type. The latter is exemplified by the segmentary societies such as the Nuer (Evans-Pritchard 1940; 1956), the Tiv (Bohannon 1957), and the Tallensi (Fortes 1945; 1949). Nadel has constructed the model for this type of society:

Think, for example, of a tribe divided into a number of sub-tribes or extended families; these all duplicate each other, both in their structure and their modes of action; each is relatively self-contained, and such relations as obtain between them (intermarriage, economic cooperation, and so forth) do not follow from their constitution (their "statutes"), but are contingent upon circumstances and outside interests. Though such segments may in fact "combine" to form the society at large, they could exist without each other and in any number; one could add to or subtract from it without affecting the working either of each segment or of the embracing group. (1951, p. 178)

Societies of this type are characterized by what Durkheim (1893) called "mechanical solidarity." Thus it is that the Tiv of Nigeria, who number perhaps a million, can still be described as a small-scale society.

The question arises as to whether such peoples as the Tiv are a "society" at all. This depends on the criteria we use to define society. There are a large number of possible criteria, such as common language, descent, origin, religion, political allegiance, and so forth. Most anthropologists would accept the criterion of the widest effective political group, i.e., the group that could employ force against outsiders and that effectively prevents or limits the use of force within the group (see Nadel 1951, p. 183-188; Mair 1962, part 1; Schapera 1956, chapter 1).

One other type of small group frequently studied by anthropologists and sociologists is the village community. Redfield has stressed the following as characteristics of the small community: distinctiveness both from the observer's and the inhabitant's point of view, smallness, homogeneity, and "all providing self-sufficiency" (1955, p. 4). Yet it is obvious that a village community is only what Kroeber ([1923] 1948, p. 284) called a "part-society." It is less self-sufficient than either the island society or the segmentary society, although, of course, there are considerable variations in this respect. [See VILLAGE.]

Scale and roles. The number of roles to be played in any society is to some extent dependent on its size, but size alone in terms of population does not mean that there are large numbers of dif-

ferent roles to be played. This is demonstrated in the populous segmentary societies. There must also be what Durkheim termed a condensation of society, multiplying social relations among more individuals and leading to a greater division of labor (1893). While the division of labor need not take place in a large population with a simple economy (as in the segmentary societies), it is more difficult for it to occur in very small-scale societies. Such societies may have a considerable proliferation of roles in the politico-ritual sphere, but they have very little specialization of economic and technical roles. Thus it is not only in the numbers of roles but also in the kinds of roles that small-scale societies differ from those of larger scale.

In their discussion of scale, the Wilsons state: "A Bushman, we maintain, is as dependent on his fellows as an Englishman, but the Englishman depends upon many more people than does the Bushman" (1945, p. 25). Yet it is obvious that there is a difference in quality in the dependence of the Bushman and of the Englishman on his fellows. The Wilsons describe this as a difference of intensity of relations, the intensity of the Englishman's relations being "more spread out."

The word "intensity" is by no means clear. The Wilsons imply that there is some total store of intensity which is invariable for all societies, but that it has various distributions in different societies. As the range of interrelations increases, the intensity of relations with near neighbors, etc., decreases. This is very difficult to prove, because we are not sure how intensity is to be measured. One could argue that the intensity of relations between spouses in London was greater (Bott 1957) than in a society in which, as in that of the Bushmen, there was greater dependence on neighbors and kinsmen. We can carry the analysis further than this by looking more closely at the nature of the roles themselves. Not only are there fewer roles in a small-scale society but also, because of the smallness of the total social field, many roles are played by relatively few individuals. It is a commonplace in anthropological studies of small communities that economic, political, religious, and kinship systems are very often congruent or nearly so. The same individuals are brought into contact over and over again in various activities. "Different types of primary groups tend to coincide or overlap in large measure" (Firth 1951, p. 47). Relationships are what Gluckman calls "multiplex," in that "nearly every social relationship serves many interests" (1955, pp. 18-19). Parsons (1939; 1951) has characterized such role relationships, which he terms "particularistic," as being affec-

tively charged. There are strong positive or negative attitudes between individuals involved in them. They also extend over a considerable time span. Such roles are usually ascriptive. The standards of judgment in the role depend on who the individual is rather than what he does. They are characterized by personal relations.

Particularistic relationships can be contrasted with impersonal relations. Parsons calls these "universalistic" because they are based on more or less fixed standards and criteria. The incumbent of a universalistic role treats all others with whom he comes in contact in terms of universal categories. A shopkeeper should treat all his customers alike, and a doctor all his patients. The roles are functionally specific in definition. The relationship, at least ideally in terms of a model, is affectively neutral. It also has a very limited time span, even though it may be repeated at intervals. The standards of judgment are based on criteria of achievement—what an individual does rather than who he is. Performance and efficiency, not ascribed hereditary qualities, are relevant. These are polar models, and it is obvious that both sets of features are characteristic of most role relationships which could be placed along a continuum. Role relationships change over time. As the same individuals continue to interact in the same roles, the relationship moves from the impersonal toward the personal pole.

In a small-scale society where the total social field is small, relationships tend to be personal. Who a man is matters very much more than what he does. For example, in business, professions, and government, family connections and friendships determine positive or negative attitudes; these are not based on role performances as shop assistants, doctors, and clerks. Occupational roles become diffuse when they have to be looked at in terms of kinship connections and influence in other spheres of activity. [See *ROLE*.]

**Values and alternatives.** An actor playing a role based on impersonal criteria is using one set of values, e.g., whether a customer owes one money; whereas the same actor acting in a personal frame uses another set of values, e.g., whether one's kinsman needs the money one loaned him. In a small-scale society one's customer is likely to be one's kinsman.

There is another sense in which the general values of a society may be related to its scale. Where kinship, economic, political, religious, and other systems tend to be coincident or nearly so, there may be a greater consistency of values than in a large-scale society, where there may be differ-



ent values for individuals acting in different situations or even for whole sets of individuals engaged in very different sorts of life (Redfield 1941). In a small-scale society anonymity is impossible. In a large-scale society, particularly in an urban setting, it is possible by moving, by changing one's job, name, or style of life. As there are more kinds of jobs and ways of life in a large-scale society, so there are more alternatives for the individual. In a small-scale society choice is limited, alternatives are few, and the choice of an individual may have considerable effects throughout the social structure. Degrees of social cohesiveness will vary from society to society even when we are only dealing with small territories. A society like that of Mauritius (Benedict 1961), with its different ethnic elements, each having its own religion and style of life, is less cohesive than the kingdom of Tonga. Yet both societies are rapidly affected by any major decisions or changes. A strike at an industrial plant in the United States has little immediate effect on most Americans; a strike at a sugar mill in Mauritius has very serious effects throughout the island, not only economically but also politically. Nearly everyone would be affected. Decisions in the economic, political, and legal fields have a pervasiveness in small-scale societies which they lack in societies of larger scale. Again, this is because people are connected with each other in so many different ways in a small-scale society. Similarly, natural disasters are more devastating in small-scale societies because their effects are so pervasive, not just on the physical environment but throughout the social fabric as well (Spillius 1957).

**Magico-religious practices.** The worship of local saints, deities, or ancestors is a characteristic of many small-scale societies and has been shown by anthropologists to be intimately bound up with personal relations within the community. This form of worship carries personal role relationships to a supernatural plane. Ancestors are an extension in time of the kin group and are intimately concerned with its welfare. Neglect of ancestors can bring misfortune in a way analogous to neglect of living relatives.

Social anthropologists following Durkheim (1912) have shown how such beliefs reflect and symbolize the solidarity of local groups, kin groups, or the whole society itself. Yet such beliefs may only symbolize the cohesion of small groups within a society of any size.

Ancestor worship is found in China as well as in small-scale societies. The notion of a high god is found in both large-scale and small-scale societies. Magico-religious practices may be an index of

the homogeneity of a society but not necessarily of its scale. It may be that very small-scale simple societies are more likely to be characterized by a single set of religious beliefs, but many small-scale societies (e.g., Guyana, Fiji, Mauritius) are remarkable for the diversity of religious beliefs found within them.

It is often assumed that secularization accompanies the growth of impersonal relationships (e.g., Redfield 1941; Wilson & Wilson 1945), but it is by no means clear how beliefs in witchcraft, sorcery, ancestors, or local deities are related to the scale of a society or whether and in what manner they inhibit the development of impersonal role relationships. The expectations that such beliefs would rapidly disappear with the spread of education and scientific knowledge have been frequently disappointed. [See MAGIC; RELIGION.]

**Jural relations.** Maine's distinction between status and contract and his famous dictum that societies have progressively moved from the former to the latter have relevance in any discussion of the sociological aspects of smallness (1861). Status in Maine's sense refers to personal role relationships. These depend chiefly on birth and who a man is. A number of anthropologists have pointed out (e.g., Bohannan 1957) that procedure in native courts is often devoted to finding out who the litigants are in terms of descent and affiliation rather than in finding out what occurred, for it is believed that an individual's behavior cannot be judged apart from who he is. The important thing is to restore good social relationships among all parties concerned (including the judges), not to conform to an impersonal law. Because of the multiple connections between litigants, lawyers, and judges in small-scale societies, there are often difficulties in applying impersonal law.

Contractual relations pose similar problems. In a contractual relation involving the payment of money, the only relevant question is whether or not the money is owed. It does not matter whether the creditor needs the money or whether the debtor can afford to pay. Where one is doing business with one's relatives, friends, and neighbors, it is difficult to apply impersonal standards. Questions of need and ability to pay may take precedence over the strict terms of the contract (see Parsons 1939). A shopkeeper having close personal ties with his clientele will find it very difficult to be an impersonal creditor, and this may well lead him into bankruptcy. It is common in many small communities for the shopkeeper, who is the principal manipulator of creditor-debtor relationships, to be an outsider, of a different ethnic or religious origin

from his clients. Thus he is not so closely connected to them by kinship and friendship ties, which enables him to be a more impersonal creditor and hence a more successful businessman (Benedict 1964, p. 344).

**Political structure.** Apart from external political factors, there are sociological factors affecting the political structure of small-scale societies. Foremost is the ubiquitousness of government. In a small-scale society one cannot progress very far up any occupational or prestige ladder without running into government. Government is an active party to nearly every sizable enterprise, not only officially but also because of the multistranded personal networks connecting the members of a small society to each other. In many technologically underdeveloped societies there are small elites marked off from the rest of the population by either ethnic criteria or class barriers. They often have control of the wealth and technical skill (including education) of the society. They usually have a large number of dependents or clients attached to them and are very often able to control the internal political machinery of the society. In small-scale societies the elite must necessarily be small. Opportunities for upward mobility are limited and more easily controlled by those in power—again because the social field is smaller. Obviously the homogeneity or heterogeneity of a society is important in this respect. Where factions form they are not apt to be simply on the basis of political issues, but extend to numerous social relationships. Where there are ethnic, religious, or linguistic differences, social cleavages may become even wider and more irreconcilable (e.g., Mayer 1963, chapter 12). Whereas in a large-scale society political relationships are only partial relationships, in a small-scale society they are much more inclusive. Closely knit family organization, personal ties within the community, traditional bonds of clientage or servitude, color bars, and so forth all militate against social mobility, whether in the political or the economic sphere.

The form of government as related to scale also deserves attention. A small governing elite is, after all, not a peculiarity of small countries. Literal democracy in the sense that everyone has a direct say in government can exist only in very small communities where everyone can meet and discuss a problem until agreement is reached (Mair 1961, p. 2). But even in small territories, with the exception of some of the very smallest islands, this is impossible. Representative government takes the place of democracy in its pure form. Theoretically, a small society with an informed electorate should

operate a representative democracy very well; but the multiplex ties connecting leaders to their followers mean that cliques or factions form which are based not on issues of policy but on ties of kinship and patronage. It is at least questionable whether very small-scale societies can successfully operate a party system. The small elites with multiplex relationships among their members may make it difficult for an opposition to develop. In a large-scale society politicians out of power can take leading roles in industry, local government, or the professions, but in small-scale societies there are relatively few such roles, and those that exist tend to be integrated with the governing elite.

For similar reasons small-scale societies often experience difficulties in developing an impersonal civil service. Civil servants form part of the governing elite. Their actions are apt to be seen as either supporting or opposing certain politicians. The network of multiplex personal relations makes it exceedingly difficult for an individual to play the role of a neutral civil servant. He is apt to be forced into partisanship (see Singham's article in Benedict 1967).

**Economic organization.** Large-scale or even medium-scale economic operations require functionally specific roles. If such enterprises are to be efficient and competitive they must be based on impersonal criteria of performance and achievement and not on an individual's personal ties with other individuals in the enterprise (United Nations 1955, p. 20). The affective component must be as neutral as possible or rational choices in terms of the efficiency of the enterprise become very difficult. Nevertheless, studies in industrial sociology have shown that where attempts are made to treat individuals as instruments rather than as individuals loss of efficiency often occurs and personal elements may have to be reintroduced (e.g., Roethlisberger & Dickson 1939). In large-scale societies there are always possibilities of bringing in outsiders. The social field is large enough to permit this. In small-scale societies outsiders must be imported from another society. Are there societies which are too small for impersonal criteria to prevail, and just how small is too small? Unfortunately, there do not seem to be any very clear-cut answers to these questions, although research could be designed which would give a more precise definition of the problem than we have now. For instance, a valuable question might be: What is the effect of the scale of a society on the specialization of roles within it?

Firth has mentioned (1951, p. 47) that there is less room for specialization of roles in a small-



scale society. This is particularly noticeable in occupational roles. Even should specialist techniques be acquired, there is simply not enough work for an individual to earn his living by his specialization alone. This situation has serious implications for economic development, for it means that a specialist must be trained and paid for performing only a very few services each year, or the society must import him at considerable expense and loss of time and at competitive prices when he is needed. Alternatively, the specialist must be a jack of all trades with the possibility that he may be master of none.

There remain questions about the type of economic enterprise and its relation to personal-impersonal pattern variables. The economies of small territories present few alternatives. They are almost invariably limited to producing a very few commodities which are exported to a limited number of markets. The domestic market is small, and local enterprise can rarely compete with cheap imports from mass producers elsewhere. Isolation from markets often aggravates the inefficiencies of personal role relationships. As the Wilsons have pointed out (1945, p. 25), there is a sense in which the scale of a society increases the more it is in contact with other societies. We need only compare Luxembourg with Mauritius, which is more than twice as populous, to see the cogency of this point. Unquestionably, the factor of development is significant. The underdeveloped countries, even the large ones, are socially characterized by personal role relationships. Their small elites reach into all facets of social existence. It can only be predicted that they have a larger potential social field for developing impersonal role relationships than the small isolated territories.

Industrialization appears to be most dependent on impersonal role relationships, while cottage industries, agriculture, and commerce seem to be less so. But some types of enterprise may thrive on a personal basis. Family businesses may succeed in risky enterprises simply because family members are not tied to the enterprise merely on the basis of a wage that they draw, but also by multiplex personal obligations, which may tide them over a period of economic difficulties. Whether a whole territory, even a small one, can succeed on such a basis is doubtful. A strong network of personal relationships or great social cohesion does not mean social harmony or common goals. The affectivity of such roles can be negative as well as positive. The intense factionalism of small communities has been observed repeatedly (Firth et al. 1957). Such

communities are no more cohesive and harmonious than groups found in large-scale societies.

What relation exists between small-scale societies defined in term of role relationships and small territories defined in terms of area and population? It is clear that small territories tend to be characterized by small-scale societies simply because of their relatively small populations. It is also clear that there are other factors involved, such as geographical isolation and economic and technological development. Luxembourg is a small territory, yet economically it exhibits large-scale characteristics as a result of its integration with its neighbors in the Common Market. Politically, however, it retains many characteristics of a small-scale society. Where there is greater physical isolation, as in the island societies of the Caribbean, the Pacific, or the Indian Ocean, the features of the small-scale society are even more marked. This is surely one of the sociological reasons attending the difficulties of federation in such areas as Indonesia, Malaysia, and the West Indies.

BURTON BENEDICT

[See also ETHNIC GROUPS; SOCIAL STRUCTURE; SOCIETAL ANALYSIS.]

#### BIBLIOGRAPHY

- BELL, NORMAN W.; and VOGEL, EZRA F. (editors) 1960 *A Modern Introduction to the Family*. London: Routledge; Glencoe, Ill.: Free Press.
- BENEDICT, BURTON 1961 *Indians in a Plural Society: A Report on Mauritius*. Great Britain, Colonial Office, Colonial Research Studies, No. 34. London: H.M. Stationery Office.
- BENEDICT, BURTON 1964 *Capital, Saving and Credit Among Mauritian Indians*. Pages 330-346 in R. W. Firth and B. S. Yamey (editors), *Capital Saving and Credit in Peasant Societies*. London: Allen & Unwin; Chicago: Aldine.
- BENEDICT, BURTON (editor) 1967 *Problems of Small Territories*. London: Athlone. → See especially the article by A. W. Singham, "Legislative-Executive Relations in Small Territories."
- BOHANNAN, PAUL 1957 *Justice and Judgment Among the Tiv*. Published for the International African Institute. Oxford Univ. Press.
- BOTT, ELIZABETH 1957 *Family and Social Network: Roles, Norms, and External Relationships in Ordinary Urban Families*. London: Tavistock.
- DURKHEIM, ÉMILE (1893) 1960 *The Division of Labor in Society*. Glencoe, Ill.: Free Press. → First published as *De la division du travail social*.
- DURKHEIM, ÉMILE (1912) 1954 *The Elementary Forms of the Religious Life*. London: Allen & Unwin; New York: Macmillan. → First published in French. A paperback edition was published in 1961 by Collier.
- EVANS-PRITCHARD, E. E. (1940) 1963 *The Nuer: A Description of the Modes of Livelihood and Political Institutions of a Nilotic People*. Oxford: Clarendon.

- EVANS-PRITCHARD, E. E. (1951a) 1956 *Social Anthropology*. London: Cohen & West; Glencoe, Ill.: Free Press.
- EVANS-PRITCHARD, E. E. 1951b *Kinship and Marriage Among the Nuer*. Oxford Univ. Press.
- EVANS-PRITCHARD, E. E. (1956) 1962 *Nuer Religion*. Oxford: Clarendon.
- FIRTH, RAYMOND (1936) 1957 *We, the Tikopia: A Sociological Study of Kinship in Polynesia*. 2d ed. London: Allen & Unwin. → A paperback edition was published in 1963 by Beacon.
- FIRTH, RAYMOND 1951 *Elements of Social Organization*. London: Watts. → A paperback edition was published in 1963 by Beacon.
- FIRTH, RAYMOND 1959 *Social Change in Tikopia: Restudy of a Polynesian Community After a Generation*. New York: Macmillan; London: Allen & Unwin.
- FIRTH, RAYMOND et al. 1957 *Factions in Indian and Overseas Indian Societies*. *British Journal of Sociology* 8:291-342.
- FORTES, MEYER 1945 *The Dynamics of Clanship Among the Tallensi*. Published for the International African Institute. Oxford Univ. Press.
- FORTES, MEYER 1949 *The Web of Kinship Among the Tallensi: The Second Part of an Analysis of the Social Structure of a Trans-Volta Tribe*. Oxford Univ. Press.
- FORTUNE, REO F. (1932) 1963 *Sorcerers of Dobu: The Social Anthropology of the Dobu Islanders of the Western Pacific*. Rev. ed. London: Routledge.
- GLUCKMAN, MAX 1955 *The Judicial Process Among the Barotse of Northern Rhodesia*. Manchester Univ. Press; Glencoe, Ill.: Free Press.
- HOMANS, GEORGE C. 1950 *The Human Group*. New York: Harcourt.
- KROEBER, A. L. (1923) 1948 *Anthropology: Race, Language, Culture, Psychology, Prehistory*. New rev. ed. New York: Harcourt.
- MAINE, HENRY J. S. (1861) 1960 *Ancient Law: Its Connection With the Early History of Society, and Its Relations to Modern Ideas*. Rev. ed. New York: Dutton; London and Toronto: Dent. → A paperback edition was published in 1963 by Beacon.
- MAIR, LUCY P. 1961 *Safeguards for Democracy*. Oxford Univ. Press.
- MAIR, LUCY P. 1962 *Primitive Government*. Baltimore: Penguin.
- MAIR, LUCY P. 1963 *New Nations*. London: Weidenfeld & Nicolson; Univ. of Chicago Press.
- MAYER, ADRIAN C. 1963 *Indians in Fiji*. Oxford Univ. Press.
- NADEL, S. F. 1951 *The Foundations of Social Anthropology*. London: Cohen & West; Glencoe, Ill.: Free Press.
- PARSONS, TALCOTT (1939) 1958 *The Professions and Social Structure*. Pages 34-49 in Talcott Parsons, *Essays in Sociological Theory, Pure and Applied*. 2d ed., rev. Glencoe, Ill.: Free Press. → First published in *Social Forces*.
- PARSONS, TALCOTT 1951 *The Social System*. Glencoe, Ill.: Free Press.
- PARSONS, TALCOTT; and SHILS, EDWARD (editors) 1951 *Toward a General Theory of Action*. Cambridge, Mass.: Harvard Univ. Press.
- REDFIELD, ROBERT 1941 *The Folk Culture of Yucatan*. Univ. of Chicago Press.
- REDFIELD, ROBERT 1955 *The Little Community: Viewpoints for the Study of a Human Whole*. Univ. of Chicago Press.
- ROETHLISBERGER, FRITZ J.; and DICKSON, WILLIAM J. 1939 *Management and the Worker: An Account of a Research Program Conducted by the Western Electric Company, Hawthorne Works, Chicago*. Cambridge, Mass.: Harvard Univ. Press.
- SCHAPERA, ISAAC 1956 *Government and Politics in Tribal Societies*. London: Watts.
- SPILLIUS, JAMES 1957 *Natural Disaster and Political Crisis in a Polynesian Society: An Exploration of Operational Research*. *Human Relations* 10:3-27; 113-125.
- UNITED NATIONS, DEPARTMENT OF ECONOMIC AND SOCIAL AFFAIRS 1955 *Processes and Problems of Industrialization in Under-developed Countries*. U. N. Doc. E/2670, ST/ECA/291. New York: United Nations.
- WHYTE, WILLIAM F. (1943) 1961 *Street Corner Society: The Social Structure of an Italian Slum*. 2d ed., enl. Univ. of Chicago Press.
- WILSON, GODFREY; and WILSON, MONICA H. 1945 *The Analysis of Social Change, Based on Observations in Central Africa*. Cambridge Univ. Press; New York: Macmillan.

## SOCIETY

The concept of society has been formulated in various ways in sociological thought. At a general level there is some agreement as to the referent of the concept, but to abstract this common meaning from the welter of different and competing meanings of the term, we must first eliminate from consideration those usages which merely adopt the term as a shorthand label for the fabric of social phenomena in general. As the historical section of this article will show, the word is often used in this all-embracing sense. Indeed, "society" is frequently used merely to refer to an encompassing network of social relationships that enclose some more specific phenomenon which is the primary object of analysis. Thus, a student of race relations may insist that interracial behavior can be understood only as a part of the organization of the larger society in which it occurs and yet may fail to provide any analytical conception of the nature of this larger whole. It is only when analysts begin to isolate the attributes of the larger whole which we term "a society" that analytical treatments of the concept begin to emerge.

**Definitions of society.** Analytical definitions usually treat a society as a relatively independent or self-sufficient population characterized by internal organization, territoriality, cultural distinctiveness, and sexual recruitment. Specific definitions vary considerably in regard to which of these elements is emphasized. For example, some stress



internal organization so strongly that they define the society as consisting in the organization and not in the population. Definitions also vary in the specific meaning given to such concepts as "self-sufficiency," "organization," and "culture." Careful formulations must develop precise apparatus for specifying the units and boundaries of societies, for describing the nature of the links between units, and for establishing the mechanisms by which these links influence social processes within society. Nevertheless, the basic concept of the inclusive, self-sufficient group remains a constant element in most concepts of society.

When the concept of society is formulated in definitive detail, it is capable of entering into sociological thought in significant ways. Various conceptions of society define the essential problems of sociological analysis and thus shape conceptions of the subject matter of sociology.

The well-known functional conception of society put forth by Aberle and his colleagues will serve to illustrate a detailed concept of society and the way it enters into a system of sociological thought. According to these analysts, "*A society is a group of human beings sharing a self-sufficient system of action which is capable of existing longer than the life-span of an individual, the group being recruited at least in part by the sexual reproduction of the members*" (Aberle et al. 1950, p. 101). The authors proceed to specify four conditions that would terminate the existence of a society: biological extinction or dispersion of the members, apathy of the members, the war of all against all, and the absorption of the society into another society. These conditions enable the authors to specify the functional requisites of a society by enumerating the mechanisms that permit a society to exist, thus specifying the meaning of the core concept of a "self-sufficient system of action." If the conditions terminating the existence of a society are to be avoided, mechanisms must develop to insure provision for sexual recruitment, provision for adequate relationship to the environment, role differentiation and the assignment of members to roles, communication, shared definitions of the world and of the goals of the society, normative regulation of means and of emotional expression, socialization, and the control of disruptive behavior. Thus, the authors have proceeded directly and logically from a definition of society to a conception of the essential concerns of societal analysis, even developing along the way some relatively specific hypotheses. For example, it is proposed that extremely rapid social change, such as revolution, tends to be accompanied by increases in the

conditions terminating the existence of a society—mortality, morbidity, apathy, force, and fraud.

It is not surprising that definitions of society are so closely articulated with conceptions of the nature and functions of sociological thought, for from the very beginning of the analytical development of the concept, social theorists have found in "society" a convenient foundation for relating their specific problems to a larger context. Thus the student interested in elucidating the nature of bureaucracy may insist that it is embedded in a larger network of organization which determines, or at least influences, the emergence of bureaucracy and the forms which it assumes and which constrains purposive attempts to shape bureaucratic development.

If society is conceived as a set of external and constraining social forces, it is natural that the concept should be put to ideological use. He who would assert that some social institution ought to change or ought to be protected from change may point to the external and constraining forces of the larger society that encompass the institution, finding in those forces both an evaluative standard and an empirical justification for that standard. The emergence of analytical concepts of society has been inextricably bound up with the development of political ideology of this type.

### History of the concept

In the Western world the concept of society as an entity distinct from the state emerged rather late. The age of reason, when philosophers began to search for secular foundations for critical analysis of existing political institutions, was one of the earliest periods when Western thinkers came to view society as something clearly prior to and outside of the state. The vehicle used to establish this differentiation was the social contract doctrine.

**The utilitarian conception of society.** In some of its forms the contract concept failed to distinguish between state and society. Thus Hobbes, wishing to insist that there is no middle ground between an organized state and the war of all against all, treated the social contract, the law of nature, and civil society as virtually identical. But the liberal thinkers of the Enlightenment wished to justify secular rational criticism of the state. In developing a critical doctrine, such thinkers as Locke began to distinguish the law of nature from the social contract that had formed the state. For Locke there is a layer of natural order guaranteed by man's interdependence and his sense of the natural rights of all. It exists prior to and outside of positive political institutions; the state is a utili-

tarian device for insuring a more efficient, less cumbersome social order by developing specialized machinery for enforcing natural law. Such a theory squarely establishes the state as a dependent sector of a larger social order.

By a similar logic the critical philosophers sought to establish analytical distinctions between society and church and to separate church and state. The church was defined by reference to its utilitarian functions for the larger society, and these functions were distinguished from the functions of the state. Thus, through the use of the concepts "civil society" and "religious society," church and state were defined as analytical functional aspects of a larger society.

Despite the incipient emergence of the conception of society as a set of interdependent organized functions, the idea of society developed during the Enlightenment was not entirely satisfactory, for the ultimate premises of argument continued to be the same premises from which Hobbes had derived the war of all against all. Enlightenment thought was founded on the concept of reason. The method of reason is analytical reduction; complex wholes must be reduced to their fundamental particles and the whole reassembled by a process of deduction from the laws governing particles. For society, the particle is the individual, and the law governing particles derives from the most essential quality of individuals, their natural reason. Each man uses his reason to rationally pursue his chosen ends. Parsons (1937) has termed this conception of society "utilitarian" and has shown that the attempt to derive social coherence and order from the faculty of reason in the individual was unsuccessful. The utilitarians could protect their Achilles heel, that is, the problem of conflicting ends, only by arbitrarily postulating such metaphysical concepts as the "natural identity of interests," "natural rights," and "the spirit of sociability." [See UTILITARIANISM, *article on SOCIOLOGICAL THOUGHT.*]

The more perceptive figures of the Enlightenment—Hume, for example—recognized the inner weakness of the utilitarian conception. However, only a sweeping new movement of thought placed the concept of society on a less atomistic foundation.

**Romanticism and organismic conceptions.** In the latter part of the eighteenth century, and especially in the period after the French Revolution, many social theorists became disillusioned with individual reason and the reductive methods of the analytical philosophers. As the philosophy of romanticism became more influential, a conservative theory of society developed which stressed the

unity of the integrated whole. Society came to be viewed as an organic growth, embodying the practical and profound wisdom of convention and tradition. Being a cumulative organic product, society has an organic unity. Abstract analytical segments cannot be separated from the whole and arbitrarily changed; to do so is to destroy the complex interdependence of the web of social life.

Again we have a concept of society as a larger whole, containing within it the standards by which particular phenomena are to be evaluated, but in this formulation the concept was put to conservative use. By assuming that the totality has an integrated character, one can effectively refute demands for arbitrary and rapid change in the parts.

The organismic conception sharpened the concept of society as a set of interdependent functions (which was already implicit in the philosophy of the Enlightenment) and drew attention to a new element, cultural tradition, as a functionally necessary part of a society. The idea of a cultural order as a constituent element of a society was developed further by August Comte in the early nineteenth century. Comte sought to synthesize Enlightened and Romantic modes of thought. Accordingly, he incorporated into his sociological system elements of classic liberal thought, such as the idea of a level of order arising from man's natural economic interdependence, and the concept of a larger society from which government derives its legitimacy. At the same time he refused to derive the larger society from individual reason and the concurrence of interests. Drawing on organismic conservatism, he found in cultural tradition the specifically collective factor in society. For Comte, the formation of any society presupposed a system of common opinions about nature and man. The Enlightenment philosophers, by destroying the normative order of the religiously based society, had loosed anarchy upon the world. Comte argued that the reformation of society required the creation of a new, scientifically based moral order. Again we see an example of the ideological use of the concept of society. Society remains a set of external and constraining forces by which we are to judge the acceptability of political ideology. The concept enters into sociological thought as a set of fundamental premises from which we are to derive the social constraints on political action.

Comte's plans for the reconstitution of society revolved about ideas for the reconstruction of religious, familial, educational, and political institutions. Embedded in his program is a contribution to the conceptual elaboration of the idea of society. Comte did not believe that a bare consensus is a



sufficient condition for organized collective life. The consensus must be organized in an institutional order that symbolizes, teaches, enforces, and implements moral ideas and rules.

The belief that society is an institutional order which embodies a fundamental set of cultural ideas was prominent in another branch of romantic thought which might be termed "idealism." Idealism, which was especially prominent in nineteenth-century German thought, stresses the cultural distinctiveness of each society. A society reflects a peculiar *Geist* or spirit that is embodied in its distinctive traditions and institutions. The spirit of a society progressively unfolds in history. Once again the ideological implications are clear: proposals for change are to be evaluated according to whether they are consonant with the cultural distinctiveness of the larger society. [See the *biographies of BOSANQUET and HEGEL*.]

**The economic conception of society.** Marx was an heir to the idealistic tradition, and the concept of society entered into his thought in a similar way. For Marx, as for the idealists, the elements of a society are closely intertwined into a complex and distinctive whole. For Marx, as for his idealistic predecessor Hegel, society is undergoing continuous transformation according to a logic of immanent development. The ideological implication is that social action is to be judged in relation to its correspondence to the immanent forces of change within society. At this point the similarity to idealism ceases, since for Marx the notion of a spirit or *Geist* is as metaphysical an explanation for the collective as the utilitarian postulate of natural order.

According to Marx, society exists in the concrete relations between social groups and not in the concepts used by philosophers to summarize these relations. The *Geist* is a mere analytical construct of the observer. The real foundations of society and the real springs of social development lie in the economic relations between men. The idea that the society outside of state and church is largely economic in character did not originate with Marx. The original utilitarian conception of society had stressed man's natural economic interdependence as the source of order which is logically prior to the state. However, Marx developed the idea of society as an economy in rich detail.

The economic conception of society starts with the assumption that man's most fundamental problem is to provide for his material needs. To do so, man must cooperate with other men by entering into relations of production. Stable relations of production constitute economic structures. Economic

structures are variable, but they generally involve two crucial phenomena: the division of men into classes and the exploitation of one class by another. Stratification and exploitation make the continuing stability of economic structures precarious, and for this reason whole complexes of compulsive apparatus develop to support the economic order. The state, law, religion, and ideology function to bring temporary stability into inherently unstable situations. Since economic structure is more basic, it can be termed the "substructure" of society; and the supporting institutions may be termed the "superstructure," since they are derivative in the sense that they are responses to the problems of economic relations. In the theory of substructure and superstructure, we see one of the first and most comprehensive theories of society as an institutional order. [See *ECONOMY AND SOCIETY*.]

**Conflict theory.** The Marxian conception of society is one of a larger set of conceptions that can be combined under the heading of "conflict theory" (Martindale 1960). The premise of conflict theory is that men are organisms, and as such they must compete for access to the resources of life. The struggle for existence does not occur between isolated individuals but between groups. In various versions of conflict theory the competing units may be families, bands, classes, nations, or races, depending on the special interests of the analyst or the stage of social development under analysis. As the conflict between groups becomes stabilized or organized or regulated, we may speak of the emergence of a structured society. Society is viewed as an organizational device for relating populations of organisms to an environment, and in this sense conflict theory may be said to adopt an ecological perspective. In conflict theory the concept of society enters into sociological thought as a means of relating social life to natural forces. In this way social scientists were able to develop a less metaphysical, more naturalistic account of society than had been produced by either Enlightened or Romantic philosophers. [See *CONFLICT*.]

To see society as a device for regulating the struggle for existence and for relating man to his environment is not necessarily to renounce interest in cultural and normative phenomena. Sumner's classic account (1906) views society as a product of antagonistic cooperation between competing organisms and groups; but the phenomenon that emerges to stabilize, organize, and regulate cooperation is a complex of customs, mores, conventions, laws, and institutions. The theory remains naturalistic in insisting that normative complexes develop gradually and naturally as a response to environ-

mental problems. They do not emanate from cultural spirits, nor are they created *de novo* from a social contract.

Conflict theory is too general a rubric to imply any particular ideological purpose. It is apparent that some forms of conflict theory are susceptible to use in support of racial or nationalistic ideologies. It is less often remarked that conflict theory was often used in the development of liberal doctrines.

The perspective of conflict theory dominated sociological thought in the latter half of the nineteenth century. Its acceptance is closely associated with the impact of the Darwinian revolution in biology and the popularity of evolutionary modes of thought. It was in connection with the use of conflict theory as an evolutionary doctrine that sociological thinkers were able to place classic liberal ideas on a naturalistic foundation.

**Emergence of the "utilitarian society."** Nineteenth-century evolutionary theory was manifold, and its devotees worked out developmental sequences for every institutional sphere of society. Insofar as the evolutionary school had any common theory of the development of total societies, it was the idea of movement toward the development of larger and more inclusive wholes. By processes of consolidation, conquest, incorporation, and differentiation, societies come to increase in scale and complexity. As this occurs, unregulated conflict between small fractionated groups becomes less important, and the regulation of internal process becomes more important. Further, as the relations with the environment and other societies become more stabilized and the larger society becomes more consolidated, new forms of social organization become possible. Social organization can be built upon processes of free discussion, free exchange, and the pursuit of individual interests. The inflexibility of the "cake of custom" and rigid military organization becomes nonadaptive; only a looser framework of organization can improve the adaptation of society to the environment by unleashing the forces of creativity and innovation.

In this brand of evolutionary theory we see the re-emergence of the concept of the "utilitarian society." However, in nineteenth-century thought the concept was no longer used as an analytical definition of the nature of any society, but as an increasingly accurate description of a historically emergent form of society.

Not all of the social analysts writing at the end of the nineteenth century viewed the emergence of the utilitarian society with equanimity. According to some analysts, the breakdown of old forms of

organization meant the loss of what had once provided society with integration, coherence, and meaning. The utilitarian society, founded upon the industrial revolution, the capitalist system, and the market mentality, fails to provide for an ethical standard outside of the individual or a viable source of social cohesion.

In 1887 Tönnies incorporated this type of perspective into his famous dichotomy between *Gemeinschaft* and *Gesellschaft*. In the *Gemeinschaft* (usually translated as "community") men are held together by communal feeling and organic ties. In the *Gesellschaft* (usually translated as "society") organic ties are replaced by artificial ties of calculating self-interest. [See COMMUNITY-SOCIETY CONTINUA.]

Thus, from varied ethical perspectives and varied ultimate premises, social theorists came to conceptualize modern society as a mere collection of individuals united only through ties of self-interest. Spencer, working in the naturalistic tradition of conflict theory, and Tönnies, drawing on a more philosophical tradition, saw modern society as a collection of wills.

**The independent reality of society.** In 1893 Durkheim severely attacked both Spencer and Tönnies and reaffirmed the reality of society as an entity. Durkheim retained an evolutionary perspective; he accepted the notion of a developmental drift from a society based on very direct ties and commonalities to a society based upon more indirect interdependencies. However, he insisted that a modern society, founded on the extensive division of labor, cannot be conceptualized as a mere collection of the wills of isolated individuals. It is not a collection of contracts based upon self-interest; it is no less organic than earlier forms of society. It remains an entity *sui generis* (Durkheim 1893).

Durkheim wrote at a time when the young discipline of sociology was struggling for establishment in an academic setting. The problem for the sociologist was to establish the independence of his field from economics, political science, and psychology. Thus, it was essential to establish society as a set of external and constraining forces, more inclusive than the economic order, outside of the state, and independent of the sum of its individual members. Durkheim adopted the strategy of stressing the independent reality of such social facts as vital rates, currents of opinion, and established conventions. Social facts require explanation at their own level. It follows that society is an entity which cannot be reduced to a set of members or a set of economic contracts between members. It exists in the complex relations and interdependencies which



unite its members into an organic whole and which in turn are reflected in a collective consciousness and a collective moral order. Persons who are drawn together in a web of interdependency interact as moral beings and create a body of collective representations of reality and regulative rules. The resulting complex regulates the utilitarian economic order and the actions of persons and is external to them.

Other analysts of the era developed similar techniques for identifying the reality of the social. Simmel found a social level in the mutual influence that interacting persons have upon each other. Mutual influence comes to have coherent forms, and thus, as people interact, they create society ([1902-1917] 1950, part 1).

Weber, although he insisted on relatively nominalistic definitions of collective entities, developed a perspective of the social order that gave a somewhat independent reality to social processes. As persons orient themselves toward each other, social relationships form; complexes of social relationships constitute a social order. Drawing upon Tönnies, Weber suggested that when actors take a rational orientation to each other, they create a *Gesellschaft*, whereas when they adopt communal solidary attitudes, they form a *Gemeinschaft*. But in either case the resulting social order is not a mere collection of wills or economic interests, since it is given stability by administrative organization, shared orientations to systems of status, and shared beliefs in the legitimacy of the order (Weber [1922] 1957, part 1).

**The social-psychological approach.** In the United States a social-psychological school emerged which found in the concept of symbolic interaction the key to the integrated treatment of society and the social person. Cooley, Mead, and others explored the development of personality and society as they emerge through interaction. Their analysis permitted a novel conceptualization of human society as a symbolically regulated process. Mead, for example, traced the process of becoming a social person. The human being comes to acquire a social personality as he learns to communicate symbolically. As he learns to adopt the perspectives of others toward himself, he is also learning to regulate his own activity symbolically by defining his self and his acts in appropriate ways. It is through participation in that complex of differentiated and interrelated roles called "society" that we develop our distinctly human capacities and identities. It is through adopting, playing, and imaginatively construing social roles that we develop social personality. Thus, self and society are intimately con-

nected through the concept of *role*. Further, the analyst may make the transition in either direction; he may stress either socialization, that is, the process by which the individual organism is socially formed, or he may stress the process by which interacting persons create and transform society. The human capacity for symbolic control over action and sympathetic understanding of the action of others makes it possible for men to innovate, to appreciate innovation, and to incorporate innovations into complex networks of symbolically regulated activity (Mead 1934).

**Society as process.** The developments of the decades when sociology emerged as a discipline can be summarized by reaffirming that the most sophisticated analysts converged on the idea that society is ultimately an organized process. If society is more than the sum of its individual participants, its reality must lie in the organized relations that emerge as men interact. The units of these relations are not people but activities. Even Durkheim, with his insistence on the independent reality of the social, recognized that the coherence of society rests on the interdependence of activities and the moral regulation created by interaction.

The new emphasis on process did not eliminate the ideological component of the concept of society. To a greater or lesser extent, all of the thinkers of the formative period saw in society a standard for evaluating the acts and programs of individuals and other social units. In some instances the ideological functions of sociological method are quite plain, as in the case of the sociological jurisprudence of Ehrlich (1913). Ehrlich saw external and constraining social reality as existing in the "living law." A spontaneous moral order develops in the associations and groups of society as men live and work cooperatively. It is this inner order of society that is the yardstick for measuring the value of the positive law of the state. Sociological jurisprudence is conceived as a corrective to conceptual jurisprudence. The latter, in its reliance on concepts and logic, is likely to become out of step with the real moral order of society and thus fail to respond to pressing social interests and problems. [See *LAW, article on THE LEGAL SYSTEM*.]

In other cases the ideological affinities of the concept of society as process are more subtle. For example, the popularity of the social-psychological approach in the United States cannot be separated from its capacity to provide a sociological foundation for modern liberalism. Classic *laissez-faire* liberalism had been founded on the utilitarian conception of society as a collection of wills. Sociological conceptions are more suited to the defense



of either conservatism or radicalism, depending on whether social reality is conceived as an irreducible obstacle or an inexorable transformative force. The problem for modern liberalism was to justify individual freedom in the context of social regulation. The concept of society as an emergent regulative process, founded on the responsible actions of social persons, is consistent with the program of modern liberalism.

### Some contemporary approaches

Despite the developments of the formative period of sociology, relatively little progress was made in defining the concept of a *society*. "Society" remained a term for designating an emergent and constraining level of reality. Until recent decades, the main purpose of analysis was to examine the nature of social reality and its modes of constraint, not to define the units and boundaries of concrete societies as entities.

**Society as a social system.** The major technical device used in recent years to attack the problem of defining a society has been the concept of a social system. A social system is an organized set of interdependent social persons, activities, or forces. It is called a system because its organization includes mechanisms for maintaining an equilibrium or some other constancy in the relations between the units. From another perspective such mechanisms can be seen as boundary-maintaining mechanisms, for systems can be isolated as separate entities only if they maintain some constancies in the face of environmental change, that is, if they maintain some boundaries vis-à-vis the environment. If every event within a system were a direct consequence of some event outside the system, it would be impossible to draw a boundary for the system; it would be, in effect, a mere unit in a larger complex. The concept of a social system seems ideally suited for use in defining a society analytically, for it contains within it the crucial concepts of "unit" and "boundary"; to define society as a special sort of social system automatically prepares the way for identification of its units and boundaries.

The idea of a social system was used by earlier thinkers, notably Spencer and Pareto, but it has been formulated in greater detail and applied to the present problem by the modern school of structural-functionalism. The paper by Aberle and his colleagues cited earlier in this article is an example of the functional mode of attack on the problem. The approach of that paper has been more fully elaborated by Levy (1952; 1966) and is similar to a definition proposed by Parsons: "A social sys-

tem . . . which meets all the essential functional prerequisites of long term persistence from within its own resources will be called a society" (1951, p. 19).

The key concept in these definitions is "self-sufficiency." All systems are self-sufficient to some degree; the definition of the term requires some mechanisms of self-maintenance. The isolation of total societies involves the search for global complexes with the highest degree of self-maintenance and the least reliance on other social systems for their requisite resources. Local communities, administrative units, institutional spheres, and bureaucratic organizations are all relatively dependent segments of a larger whole. Total societies are dependent only on other types of systems: personalities, cultural traditions, and a physical environment. [See SYSTEMS ANALYSIS.]

**Overlapping process systems.** The search for the self-sufficient society may be futile. In fact, the concept of a *society* with exclusive boundaries may be obsolete. This is not to say that it is not valid to do comparative study over very large aggregates. Indeed, one of the achievements of twentieth-century empirical social science is the thorough demonstration that society is real in the sense that social variables relate to each other differently in "societies" with different characteristics. In this sense the concept of society no longer rests on conjecture. However, at the theoretical level the search for a universally valid definition of the nature and boundaries of a society as a self-contained unit may obscure the complexity of social life.

Many of the historic and contemporary problems in the conceptual analysis of society may be clarified by viewing society as a complex of overlapping process systems. We may abstract from the concrete interaction of concrete social persons a number of types of interaction systems. Economic, religious, political, educational, and other types of activity come to cohere into partially independent systems with units, boundaries, and mechanisms of their own. These systems overlap; and when a relatively broad range of such systems cohere around a common population, we may speak of a society. There is no reason to suppose, however, that this society will be self-contained, that it will not overlap with other societies, or that its boundaries will be uniform across its constituent systems.

It would be a mistake to regard the notion of a self-sufficient society as intrinsically absurd. One might be tempted to argue that any society engages in at least some trade and cultural exchange with at least one other society, and that therefore there are no self-sufficient societies. This argument, how-



ever, misconstrues the concept of self-sufficiency. To call a society self-sufficient implies not that it is isolated but that its social system contains within itself cultural materials and role opportunities sufficient for carrying on controlled relations with an environment (Parsons 1966).

The truth or falsehood of alternative conceptions of society is not at issue here. Rather, we are contrasting alternative ways of approaching the problems of societal analysis. Thus, an approach based on the concept of social system can be criticized for failing to make several important phenomena immediately problematic. But this is very far from saying that it causes logical difficulties so severe as to prevent analysis of these phenomena altogether.

The alternative to the social system approach is to start with the concept of population. This involves making problematic not only the degree to which the activities of a population are organized into systems, but also the units and boundaries of these systems and their forms of interdependence. Thus the emergence of a bounded, unified social system is no longer assumed but becomes an object of inquiry.

The boundaries of a society are usually drawn at the outer limits of the interdependencies and commonalities that give it coherence. Special criteria are often used to define particular types of society. Thus, in the case of the nation-state the limits of political jurisdiction are taken to define the limits of the political system and, thus, of the society. In fact, if societies are viewed as complexes of overlapping process systems, it becomes natural to accept the fact that societal boundaries are irregular. From this, one is led to consider the sources and consequences of variation in the scope of systemic boundaries (for instance, by examining the consequences of the involvement of national states in international colonial development).

Another single-criterion approach to defining the boundaries of the societal population is to adopt the social system's own normatively defined concept of membership (Parsons 1966). To this it may be objected that the members of a societal population may be involved in well-organized social systems that crosscut membership boundaries. The common solution to this problem is to treat the basic unit of society as the "person-in-role," and then to allege that behavior in boundary-crossing roles is *external* to the society. One then attempts to identify the mechanisms for segregating roles and regulating members who cross societal boundaries. Again, the focus is on how such regulation is accomplished, not on the conditions under which it will emerge or the consequences of its absence.

To define society as a set of overlapping process systems also permits flexibility in the analysis of the units of society. The utilitarian conception of society was based on the notion of the isolated individual. As early as Comte, the wisdom of this approach was called into question. Comte saw the utility of beginning with more socially relevant units; for him, the unit of society was the family. Other analysts continued this tradition in their treatment of the communal aspects of society. Later, with the development of processual concepts, segments of the actions of social persons came to be used as units of analysis. Thus, Parsons has spoken of the units of social systems as actors in roles (1951, pp. 24-26).

If society is viewed as a complex of overlapping process systems, it would appear that the ultimate social unit is the act. However, it does not follow that the intermediate units of different process systems will be the same. Society does not have a single type of unit, but many: its constituent systems have units of different types, and these differences can be of considerable analytical importance. For example, many of the strains in modern societies derive from the fact that, although families are the molecular units of modern status systems, rational market and bureaucratic systems are predicated on the evaluation of specific performances in roles. In consequence, competing bases of evaluation are present in society.

The links between units may also vary from system to system. Theories of society have postulated six major types of links between units: emotional attraction, orientations of actors to each other, shared cognitive and evaluative perspectives, mutual influence or coercion, economic or functional interdependence, and common participation in an environment. There is no reason to assume that any of these types of links constitutes the one true mechanism of social coherence. On the other hand, one or another of them may be dominant in particular process systems. This possibility is itself a source of important sociological problems: we are led to ask such questions as "What is the consequence of the expansion of systems of economic interdependence beyond the boundaries of systems based upon emotional attachments?"

The notion of a set of overlapping process systems is not by itself a satisfactory model of society. It leaves open the question of the relation between a society and other social groupings. How, for instance, is a society to be distinguished from a community? The term community has been used in a variety of ways. For some, communities are locally based units of a larger society; for others, "com-



munity" refers to some aspect of society, such as its solidarity (that is, communal) or spatial components. Others, particularly in the German sociological tradition, distinguish communities as relatively solitary types of societies [see COMMUNITY-SOCIETY CONTINUA].

It is legitimate to use the term "community" to refer to both locally based units and some aspect of the larger society. The former usage is well established in English speech. The latter usage is sometimes said to be justified by the need to establish a distinction between societies and other social entities. Institutions and special-purpose associations are said not to be societies because they have no communal components. The concept of population can be used in a similar way to distinguish societies from other sets of systems of social processes, since the latter may have sets of members without having populations in the biotic sense. A society is sustained by a population. To establish the boundaries of a societal population we may adopt a definition of population quite similar to the one employed by bioecologists: A population consists of the self-perpetuating inhabitants of a territorial area. In this context the term "self-perpetuation" implies mating, and the term "inhabitant" implies relatively permanent residence. Thus, the boundaries of a population that sustains a society are established by the limits of the largest territorial area within which mating is common and residence is relatively permanent.

These criteria establish a model that, while it does not completely correspond to any empirical population, does provide a means of clearly excluding types of groups that are not societies. Many groups do not qualify as societies because membership in them is not conferred by birth. Further, members of local communities quite often cross community boundaries for the purpose of mating and establishing new residences, while members of societies do not often cross societal boundaries for these purposes—indeed, that is why they form what I have called a population.

To start with the concept of a population is not to define a society as consisting in its population. The society is not the population but the complex systems of action in which the units of the population participate.

The next problem for the analyst is to establish the boundaries of these systems of action. In a highly organized society, which closely controls the relations between the units of its population and members of other populations, it may be useful to treat only relations within the societal population as internal to the society. On the other hand, when

societal systems become very permeable to social influences that transcend population boundaries, it is more realistic to consider the society to have irregular boundaries and to overlap other societies.

Historically, conceptions of society have had ideological implications and overtones. Perhaps this conception of society as a set of overlapping process systems is not an exception, for it seeks to call attention to the problems of the emergence of larger and more inclusive networks of social organization. As man has expanded his ecological niche there has been a continuous growth of national organization that transcends the less inclusive traditional solidarities. At the same time, cosmopolitan and international organization has outrun national boundaries. The modern world is regularly upset by shock waves reverberating from local traditional cleavages through national political systems; sometimes these shock waves reach the international arena. If sociological analysis is adequately to represent the constraints imposed by this emergent global level of social reality, its analytical conceptions must not be inflexibly tied to the concept of the national boundary.

LEON H. MAYHEW

[See also COOPERATION; CULTURE; EVOLUTION; INTEGRATION; INTERNATIONAL INTEGRATION; NATURAL LAW; POLITICAL SOCIOLOGY; SOCIAL CONTRACT; SOCIAL DARWINISM; SOCIOLOGY, article on the DEVELOPMENT OF SOCIOLOGICAL THOUGHT; and the biographies of COMTE; COOLEY; DURKHEIM; HOBBS; LOCKE; MARX; MEAD; MILL; SIMMEL; SPENCER; SUMNER; TÖNNIES; WEBER, MAX.]

#### BIBLIOGRAPHY

- ABERLE, DAVID F. et al. 1950 The Functional Prerequisites of a Society. *Ethics* 60:100-111.
- DURKHEIM, ÉMILE (1893) 1960 *The Division of Labor in Society*. Glencoe, Ill.: Free Press. → First published as *De la division du travail social*.
- EHRlich, EUGEN (1913) 1936 *Fundamental Principles of the Sociology of Law*. Translated by Walter L. Moll with an introduction by Roscoe Pound. Cambridge, Mass.: Harvard Univ. Press. → First published as *Grundlegung der Soziologie des Rechts*.
- LEVY, MARION J. JR. 1952 *The Structure of Society*. Princeton Univ. Press.
- LEVY, MARION J. JR. 1966 *Modernization and the Structure of Society: A Setting for International Affairs*. 2 vols. Princeton Univ. Press.
- MARTINDALE, DON 1960 *The Nature and Types of Sociological Theory*. Boston: Houghton Mifflin.
- MEAD, GEORGE H. (1934) 1963 *Mind, Self and Society From the Standpoint of a Social Behaviorist*. Edited by Charles W. Morris. Univ. of Chicago Press. → Published posthumously.
- PARSONS, TALCOTT (1937) 1949 *The Structure of Social Action: A Study in Social Theory With Special Reference to a Group of Recent European Writers*. Glencoe, Ill.: Free Press.



- PARSONS, TALCOTT 1951 *The Social System*. Glencoe, Ill.: Free Press.
- PARSONS, TALCOTT 1966 *Societies: Evolutionary and Comparative Perspectives*. Englewood Cliffs, N.J.: Prentice-Hall.
- SIMMEL, GEORG (1902-1917) 1950 *The Sociology of Georg Simmel*. Edited and translated by Kurt H. Wolff. Glencoe, Ill.: Free Press. → First published in German.
- SUMNER, WILLIAM GRAHAM (1906) 1959 *Folkways: A Study of the Sociological Importance of Usages, Manners, Customs, Mores, and Morals*. New York: Dover.

→ A paperback edition was published in 1960 by New American Library.

- TÖNNIES, FERDINAND (1887) 1957 *Community and Society (Gemeinschaft und Gesellschaft)*. Translated and edited by Charles P. Loomis. East Lansing: Michigan State Univ. Press. → First published in German. A paperback edition was published in 1963 by Harper.
- WEBER, MAX (1922) 1957 *The Theory of Social and Economic Organization*. Edited by Talcott Parsons. Glencoe, Ill.: Free Press. → First published as Part 1 of *Wirtschaft und Gesellschaft*.